2011 NASA Statistical Engineering Symposium - Agenda At-a-Glance

	Tuesday (May 3rd)	Wednesday (May 4th)		Thursday (May 5th)	
Start	**Cascades Room**	**Track A (Cascades)**	**Track B (Azalea)**	**Track A (Cascades)**	**Track B (Azalea)**
8:00					
8:15					
8:30		**Plenary Session**		**Plenary Session**	
8:45		*Snee*		*Wilson*	
9:00					
9:15		Break		Break	
9:30		**Session 1A**	**Session 1B**	**Session 4A**	**Session 4B**
9:45		*Hutchinson*	*DeLoach*	*Simpson*	*Wilson*
10:00		*Parker*	*Silvestrini*	*Hutto*	*Volovoi*
10:15					
10:30		Break		Break	
10:45					
11:00		**Session 2A**	**Session 2B**	**Session 5A**	**Session 5B**
11:15		*Kenny*	*Womack*	*Freeman*	*Phojanamongkolkij*
11:30		*Ferson*	*White*	*Anderson-Cook*	*Brown*
11:45					
12:00		Break		Break	
12:15					
12:30	**Welcome**				
12:45	**Opening Address**	**Lucheon**		**Lucheon**	
1:00	*Hoerl*	Oak Room		Oak Room	
1:15					
1:30					
1:45	Break	Break		Break	
2:00	**Keynote Address**	**Session 3A**	**Session 3B**	**Session 6A**	**Session 6B**
2:15	*Gilmore*	*Vining*	*Guthrie*	*Reinman*	*Rhew*
2:30		*Landman*	*McCollum*	*Safie*	*Lynn*
2:45		*Borror*	*Johnson, T.*	*Vesely*	*Commo*
3:00	Break				
3:15	Leadership Panel				
3:30	***Integrating***	Break		Break	
3:45	***Critical, Statistical Thinking***				
4:00	*Moderator: Turner*	Panel Discussion		Panel Discussion	
4:15		**Statistical Research**		**Statistical Engineering**	
4:30		*Moderator: Vining*		*Moderator: Anderson-Cook*	
4:45					
5:00					
5:15	**Reception**				
5:30	**Conference Center**				
5:45					
6:00		Banquet at the Conference Center		Bus Transport to Banquet at Williamsburg Winery	

2011 NASA Statistical Engineering Symposium
Agenda

Tuesday, May 3rd

12:30 – 12:45 Welcome and Opening Remarks
Mr. Stephen Jurczyk
Deputy Director, NASA Langley Research Center

12:45 – 1:45 Opening Address
Why Statistical Engineering?
Dr. Roger Hoerl
Lead, Applied Statistics Laboratory, GE Global Research

1:45 – 2:00 Break

2:00 – 3:00 Keynote Address
Increasing Statistical Rigor in Operational Test and Evaluation
Dr. J. Michael Gilmore
Director of Operational Test and Evaluation, Office of the Secretary of Defense

3:00 – 3:15 Break

3:15 – 4:30 Senior Leadership Panel
Integrating Critical, Statistical Thinking
Mr. Clayton Turner (Moderator)
Chief Engineer, NASA Langley Research Center

Dr. L. DeWayne Cecil
Western Region Climate Services Director, National Oceanic and Atmospheric Administration

Mr. Steven Gentz
Chief Engineer, NASA Engineering and Safety Center, Marshall Space Flight Center

Dr. Edward Kraft
Chief Technologist, Arnold Engineering Development Center

Mr. Stephen Sandford
Director, Engineering Directorate, NASA Langley Research Center

5:00 – 6:00 Reception

Wednesday May 4th

8:30-9:15

Leadership – Essential for Developing the Discipline of Statistical Engineering

Ronald Snee, *Snee Consulting*

9:30 - 10:30

1A - Infusing Statistical Engineering at NASA Langley Research Center

Chair: Paul Roberts, *NASA Engineering & Safety Center*

Building Capability
Mark Hutchinson
NASA Langley Research Center

Infusing Statistical Engineering
Peter Parker
NASA Langley Research Center

1B – Research and Applications of Response Surface Methodology

Chair: Steve Helland, *NASA Glenn Research Center*

Topics in Response Surface Methodology Adequacy Assurance and Assessment
Richard DeLoach
NASA Langley Research Center

Hybrid Designs: Space Filling and Optimal Experimental Designs for Use in Studying Computer Simulation Models
Rachel Johnson Silvestrini
Naval Postgraduate School

11:00 – 12:00

2A - Uncertainty Propagation in Dynamic Models

Chair: Dr. Luis Crespo, *National Institute of Aerospace*

Failure Domain Bounding with Applications to Dynamic Systems
Sean Kenny
NASA Langley Research Center

Beyond Probability: A Pragmatic Approach to Uncertainty Quantification in Engineering
Scott Ferson
Applied Biomathematics

2B - Requirements Verification

Chair: Ken Johnson, *NASA Engineering & Safety Center*

Statistical Tolerance Bounds: Overview and Applications to Space Systems
James Womack
Aerospace Corporation

An Empirical Study of Variables Acceptance Sampling: Methods, Implementation, Testing, and Recommendations
K. Preston White
University of Virginia

12:15 – 1:45 Luncheon at Conference Center

2:00 – 3:30

3A – Statistical Collaboration with University Partners

Chair: Jim Simpson, *United States Air Force*

Opportunities for Statistical Collaboration with NASA: Some Personal Reflections
Geoff Vining
Virginia Tech

Designed Experiments in Aerospace Ground Testing: Challenges and Successes
Drew Landman
Old Dominion University

Quality Engineering: A Journal Dedicated to Quality Improvement Methods and Applications
Connie Borror
Arizona State University

3B – Measurement System Uncertainty Analysis

Chair: Mark Zabel, *Straight Line Performance Solutions*

The metRology Package in R: Tools for Statistical Metrology and Uncertainty Analysis
Will Guthrie
National Institute of Standards Technology

Modeling the Kronecker Product Covariance Structure for Canonical Correlation Analysis
Ray McCollum
Booz Allen Hamilton

Design of Experiments in Measurement System Characterization and Uncertainty
Thomas Johnson
NASA Langley Research Center

4:00 – 5:30

Panel: Statistical Research

Moderator: Geoff Vining, *Virginia Tech*

Panel: Christine Anderson-Cook (*Los Alamos National Labs*), Carolyn Morgan (*Hampton University*), Ray Rhew (*NASA Langley*)

6:00 Banquet Dinner at Conference Center, Special Guest Thomas Jefferson

Thursday May 5th

8:30 – 9:15

Overview of the NASA Engineering and Safety Center and Leveraging Limited Data: A Challenge for Statistical Engineering

Timmy Wilson, *Deputy Director of NASA Engineering and Safety Center*

9:30 – 10:30

4A – Statistical and Systems Engineering in the US Air Force Test and Evaluation Enterprise

Chair: Peter Parker, *NASA Langley Research Center*

Engineering Test Science for the Military
Jim Simpson
United States Air Force, 53rd Test Wing

Doing the Right Things: Experimenting So That Warriors Do Not
Greg Hutto
United States Air Force 46th Test Wing

4B – Next Generation Airspace Applications

Chair: Steve Velotas, *NASA Langley*

Statistical Design and Analysis of Experiments for Next Generation Air Transportation Research
Sara Wilson
NASA Langley Research Center

System Safety and Reliability Modeling for the Next Generation of Air Transportation
Vitali Volovoi
Georgia Tech

11:00 – 12:00

5A – Improving Reliability Modeling for Engineers

Chair: Geoff Vining, *Virginia Tech*

Accelerated Life Testing: Tutorial with Applications in NASA and the DoD
Laura Freeman
Institute for Defense Analysis

Modeling the Reliability of Complex Systems with Multiple Data Sources: A Case Study on Making Statistical Tools Accessible to Engineers
Christine Anderson-Cook
Los Alamos National Laboratory

5B – Statistical Contributions to Research and Policy in Climate Observational Systems and Modeling

Chair: Ray Rhew, *NASA Langley*

Developing a Measurement System Uncertainty Framework for Earth Observing Satellites
Nipa Phojanamongkolkij
NASA Langley Research Center

Responding to Climate Variability and Change: A Rapid Prototype For Assessing Impacts of Uncertainty in Climate Observations and Model Projections on Decision Support
Douglas Brown
Booz Allen Hamilton

12:15 – 1:45 Luncheon at Conference Center

2:00 – 3:30

6A – Design for Variation & Reliability and Maintainability Engineering

Chair: Tim Adams, *NASA Kennedy Space Center*

Design for Variation at Pratt & Whitney
Grant Reinman
Pratt & Whitney

A Statistical Approach for Life Limits of Space Shuttle Main Engine Components
Fayssal Safie
NASA Safety Center

Applications of Bayesian Statistical Analyses in Determining the Number of Demonstration Tests to Conduct and in Monitoring Reliability Growth
Bill Vesely
NASA Headquarters

6B – Measurement System Characterization with Applications in Aeronautics and Exploration

Chair: Mark Schoenenberger, *NASA Langley*

Advancements in Aeronautics Measurement System Characterization
Ray Rhew
NASA Langley Research Center

Perspectives on Planetary Entry, Descent, and Landing Research
Sean Commo
NASA Langley Research Center

Thermal and Pressure Characterization of a Wind Tunnel Force Balance using the Single Vector System
Chris Lynn
NASA Langley Research Center

4:00 – 5:30

Panel: Statistical Engineering: What, Why and How

Moderator: Christine Anderson-Cook, *Los Alamos National Laboratory*

Panel: Ronald Snee (*Snee Consulting*), Geoff Vining (*Virginia Tech*), Mark Zabel (*Straight Line Performance Solutions*)

6:00 Bus Transportation to Banquet Dinner at the Williamsburg Winery (Dinner at 6:30)

Why Statistical Engineering?

NASA Conference on Statistical Engineering
Williamsburg, VA
May 3rd, 2011

Roger Hoerl, GE Global Research

With Considerable Input From Ronald D. Snee

Today's Agenda

- Motivation
- Today's Realities
- My "Blinding Light of the Obvious"
- Statistical Engineering Definition
  - Giving a Name to Something Too Often Done in the Shadows
- Examples
- Why Statistical Engineering Needs to be a Formal Discipline
  - Versus Done on an Ad-Hoc Basis
- Summary

My Motivation

- Internship at DuPont
- Early career at Hercules Chemical Co.
- Wake up call at Scott Paper Co.
- Career at GE

For me, this is personal

Today's Realities

The statistics discipline has much to be proud of and excited about:

- Hal Varian (chief economist at Google): "I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding."
- More interdisciplinary academic research being conducted
- Increased enrollments in service courses at many universities

At the same time, there is an ongoing concern that statisticians are not fully utilized within their organizations:

- Statisticians often are not sure how to take more initiative to get involved in the large, complex, unstructured "mission critical" projects in their organizations
- Management may see that statisticians can make larger contributions than passive consulting, but are not sure how to properly deploy them
- Professional colleagues may not understand how to engage statisticians in broader roles as collaborators, as opposed to going to them for support on narrow technical questions

Statisticians have much to offer: how do we unlock their potential?

My "Blinding Light of the Obvious"

Susan Hockfield – MIT President:

Around the dawn of the 20th century, physicists discovered the basic building blocks of the universe; a "parts list", if you will. Engineers said "we can build something from this list," and produced the electronics revolution, and subsequently the computer revolution.

More recently, biologists have discovered and mapped the basic "parts list" of life – the human genome. Engineers have said "we can build something from this list," and are producing a revolution in personalized medicine.

Loosely quoted from January, 2010 seminar at GE Global Research

Application to Statistical Science

Two important questions we must answer:

Who is building something meaningful from the statistical science "parts list" of methods?

What are the implications of stopping at developing the parts list – the methods, and not building something of interest to society from them?

Can statisticians be thought leaders in addition to being "tools guys"?

Some Current Challenges

Some currently unsolved statistics problems:

- Ensuring that statistical projects have high impact
- How to attack big, complex, unstructured problems
  - Problems that do not "correspond to an identifiable textbook chapter" (Meng, 2009)
- The need to integrate the principles of statistical thinking with the application of statistical methods and tools
- Providing opportunities for statisticians to demonstrate true leadership to their organizations, rather than only passive consulting services

An opportunity to build something new from the parts list?

A Conjecture

Scientists, engineers, statisticians and other professionals have been building meaningful new things from the statistical science parts list of tools for some time, to society's benefit. However:

- This has typically been done on an ad-hoc basis with little or no underlying theory documented to provide guidance to others.
- Applications have generally been "one offs", requiring the wheel to be reinvented each time.
- This has significantly slowed progress, and missed opportunities to benefit society.

Not a new idea, but perhaps a new discipline

Statistical Engineering Fundamentals

Science:

- Systematic study and advancement of the facts and general laws of the physical world

Engineering:

- Study of how to best utilize scientific and mathematical principles for the benefit of humankind

Note: while scientists study and advance the fundamental laws of nature, engineers study how existing laws and principles could be put to better use, e.g., the IBM Computer "Watson"

- Engineers develop theory - How existing science can be better utilized - What works, what doesn't, and why
- The development and use of theory is the key differentiator between an "engineer" and a "practitioner"

Theory: A coherent group of general propositions used to explain a class of phenomena

- Note: there is no mention of mathematics in this or other common definitions of theory!

Terminology is important, and needs to be precise

Statistical Engineering Definition

Statistical engineering:

- The study of how to best utilize statistical concepts, methods, and tools and integrate them with information technology and other relevant sciences to generate improved results (Hoerl and Snee 2010a)
- In other words, trying to build something meaningful from the statistical science parts list

Notes

For this to be a true engineering discipline as opposed to just a sexier term for applied statistics, there must be a dynamic theoretical foundation based on rigorous research, just as there is for electrical engineering, mechanical engineering, and so on

This definition does not refer to application of statistics to engineering

- Statistical engineering can be applied to improving anything

This is a different definition than that used by Eisenhart, who we believe was the first to use this term in 1950

This definition is consistent with dictionary definitions of engineering

Elaboration of Definition

An important viewpoint to keep in mind:

The issues we raise above have nothing to do with the old distinction between applied statistics and theoretical statistics. The traditional viewpoint equates statistical theory with mathematics and thence with intellectual depth and rigor, but this misrepresents the notion of theory.

We agree with the viewpoint that David Cox expressed at the 2002 NSF Workshop on the Future of Statistics that "theory is primarily conceptual," rather than mathematical. (Lindsay et al. 2004)

- Statistical engineering is different from traditional applied statistics
- Theory is conceptual, not necessarily mathematical

Applied Statistics or Statistical Engineering?

Three Work Environments
Experienced by Roger Hoerl Early in his Career

Work Environment	Work Description	Comments
DuPont Engineering– Summer Intern 1981, 1982	Use statistics methods to address design and analysis problems presented to him	• Applied methods learned in graduate school. • Employer, co-workers and clients were pleased
Hercules 1983-87	Identify opportunities to apply statistical methods to important design and analysis problems	• Deepened understanding of how to apply methods learned in graduate school. • Employer, co-workers and clients were very pleased
Scott Paper 1987-1995	Deploy Statistical Process Control across the company	• Not covered in graduate school • Not covered in SPC books • Other skills beyond statistics were needed • Several concepts, methods and tools needed to be integrated

Example of Statistical Engineering:

Lean Six Sigma and the DMAIC Framework

Lean Six Sigma (LSS) is an obvious example of statistical engineering – building something from the parts list of tools. I believe that this is one reason for its prolonged success. LSS has been deployed broadly not only within the private sector, but within several branches of the US armed forces.

- LSS has not invented any new tools (no new statistical science), but has achieved much greater results from existing tools
  - The tools have been more effectively integrated and linked through the DMAIC model, for example
- LSS has developed dynamically through cycles of application of the scientific method – theory and experimentation
  - Addition of the Define stage, integration of Lean approaches, etc.
- LSS development has been based on a solid theoretical foundation in continuous improvement, e.g.:
  - Pareto principle (focus on a few key process drivers)
  - Use of project-by-project improvement (Juran)
  - Utilization of small (4-6 people) project teams (Useem 2006)
  - Use of tools with sound theoretical bases; DOE, regression, SPC, etc.

A result of "the study of how to best utilize statistical concepts, methods, …"

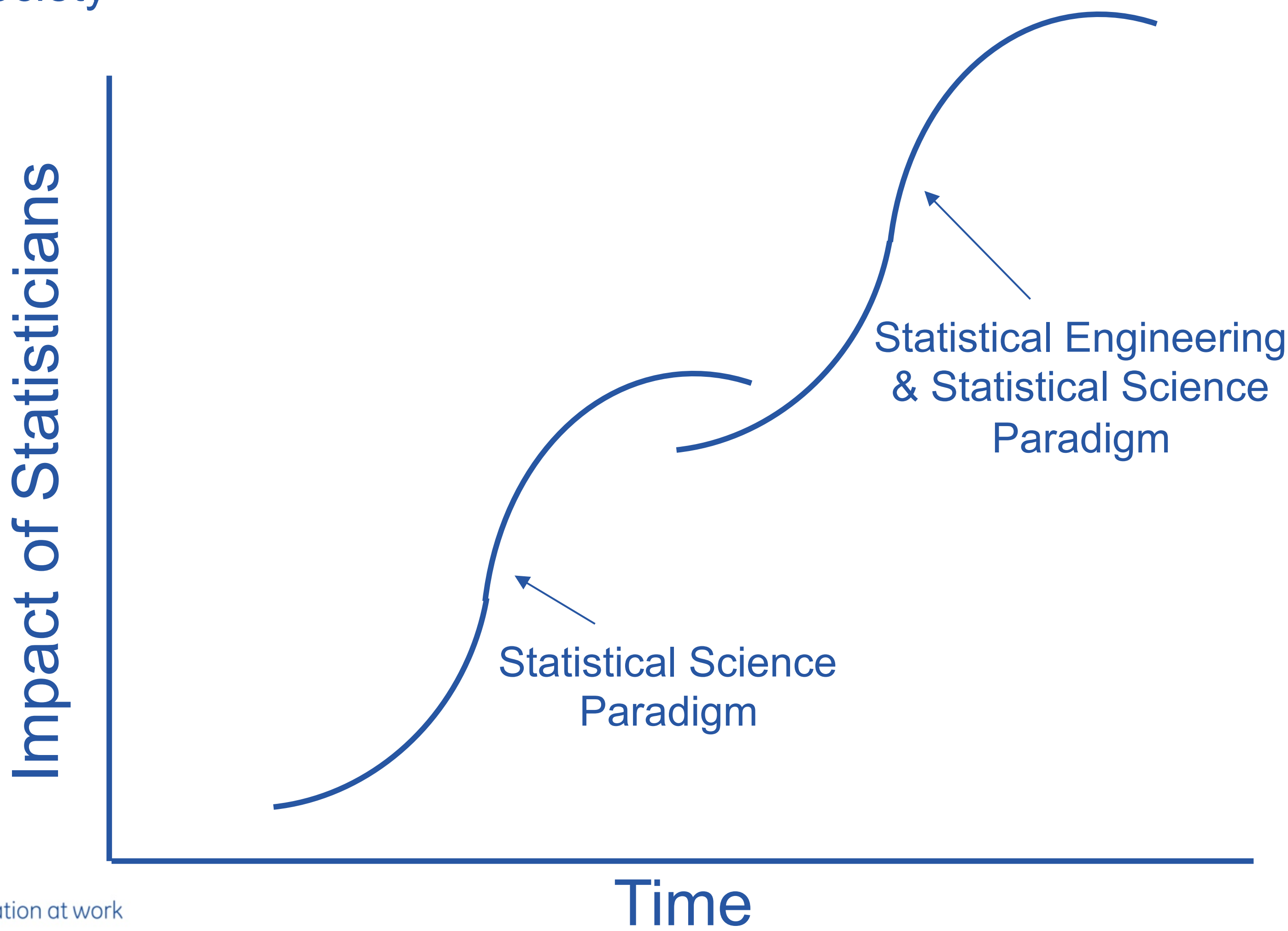
A Personal Example

- Problem: Develop a financial "default predictor" for GE
  - Big, unsolved problem!
  - Challenges of trading within a multi-billion dollar portfolio
  - No commonly accepted definition of default
  - Limited internal data – no set of "universal data" exists
  - No defined measure of success
- Approach:
  - Cross-functional team needed (statistics, OR, quantitative finance, business expertise), spread between in upstate New York, Bangalore, and Stamford, CT
  - Developed definition of default, and metrics to document success and failure
  - Data obtained externally – needed to merge disparate data sources
    - Set up direct feed from Wall Street
  - Final prediction methodology utilized:
    - Publicly available default predictor as an input
    - Smoothing algorithms, CART, simulation, Markov Chains, and censored data methods from reliability
  - Awarded US Patent for *system* – not for *algorithm*

Not at textbook problem – or solution

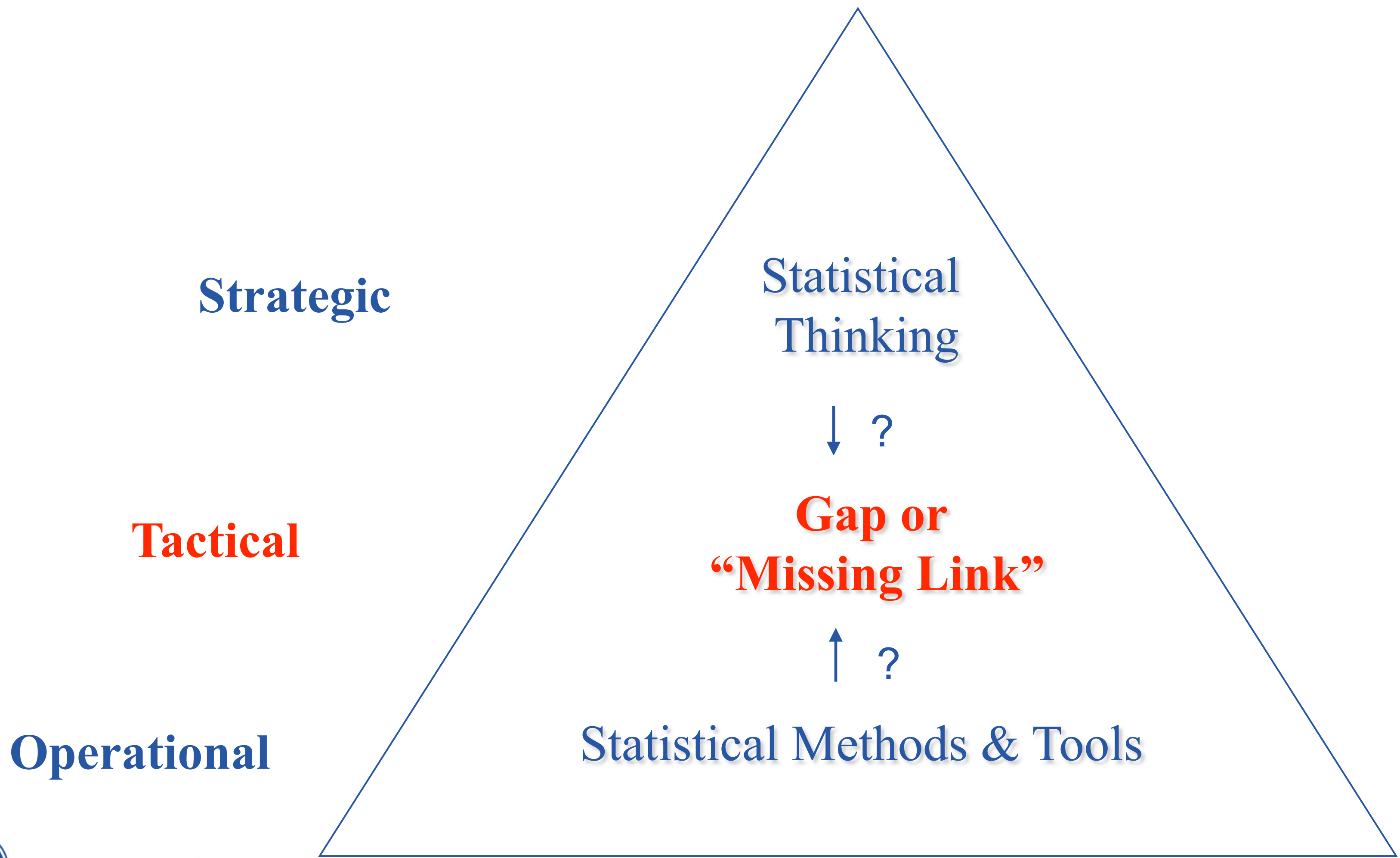
Potential Impact of Statistical Engineering

I believe that a balanced approach involving both statistical science and statistical engineering would enable us to have a much broader impact on society



Potential Impact of Statistical Engineering

Many have noted a gap between the concepts of statistical thinking and the application of statistical tools and methods



imagination at work

Statistical Thinking Definition

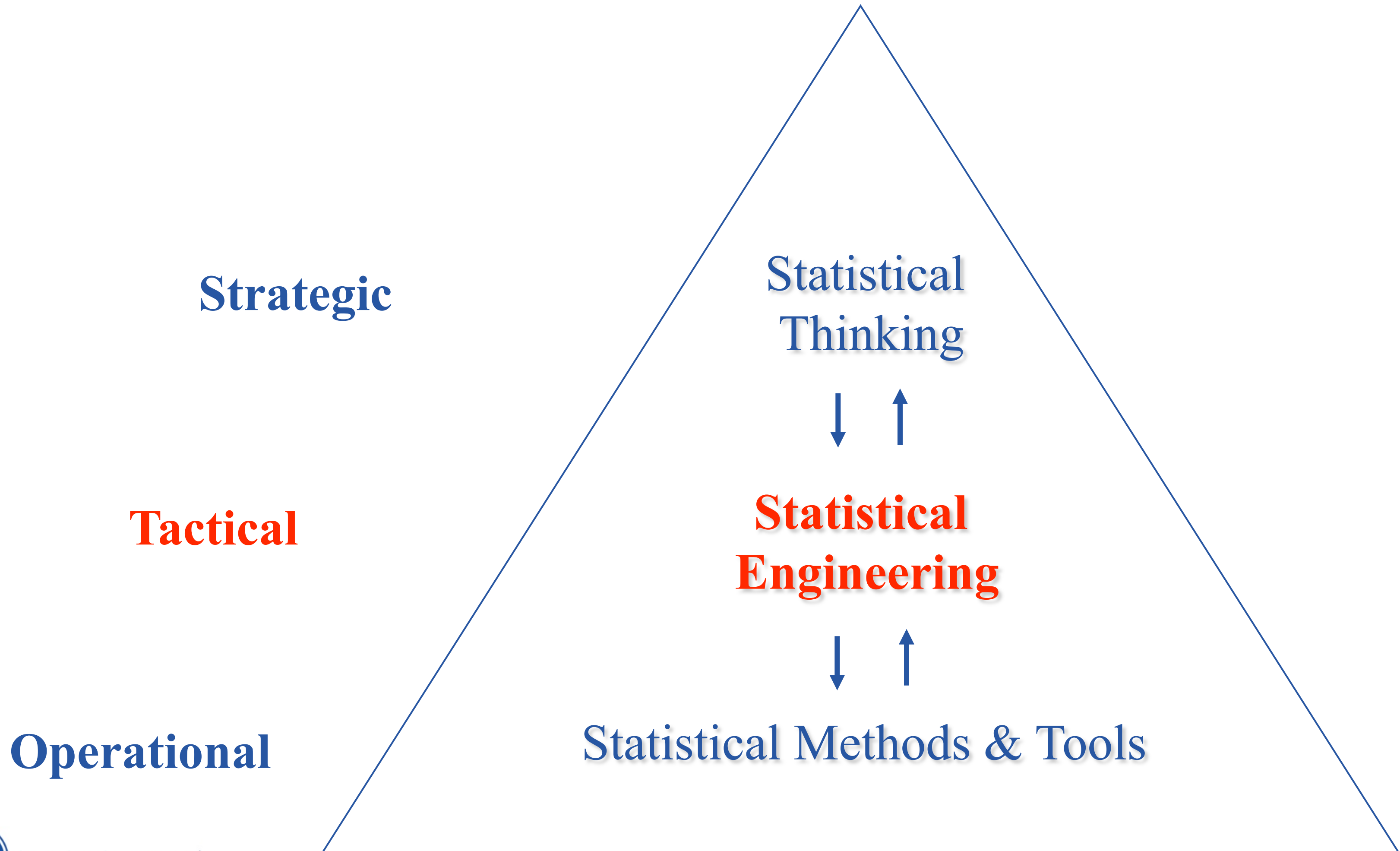
Statistical Thinking is a philosophy of learning and action based on these fundamental principles:

- All work occurs in a system of interconnected processes
- Variation exists in all processes
- Understanding and reducing variation are keys to success

ASQ Glossary and Tables for SQC, 1996

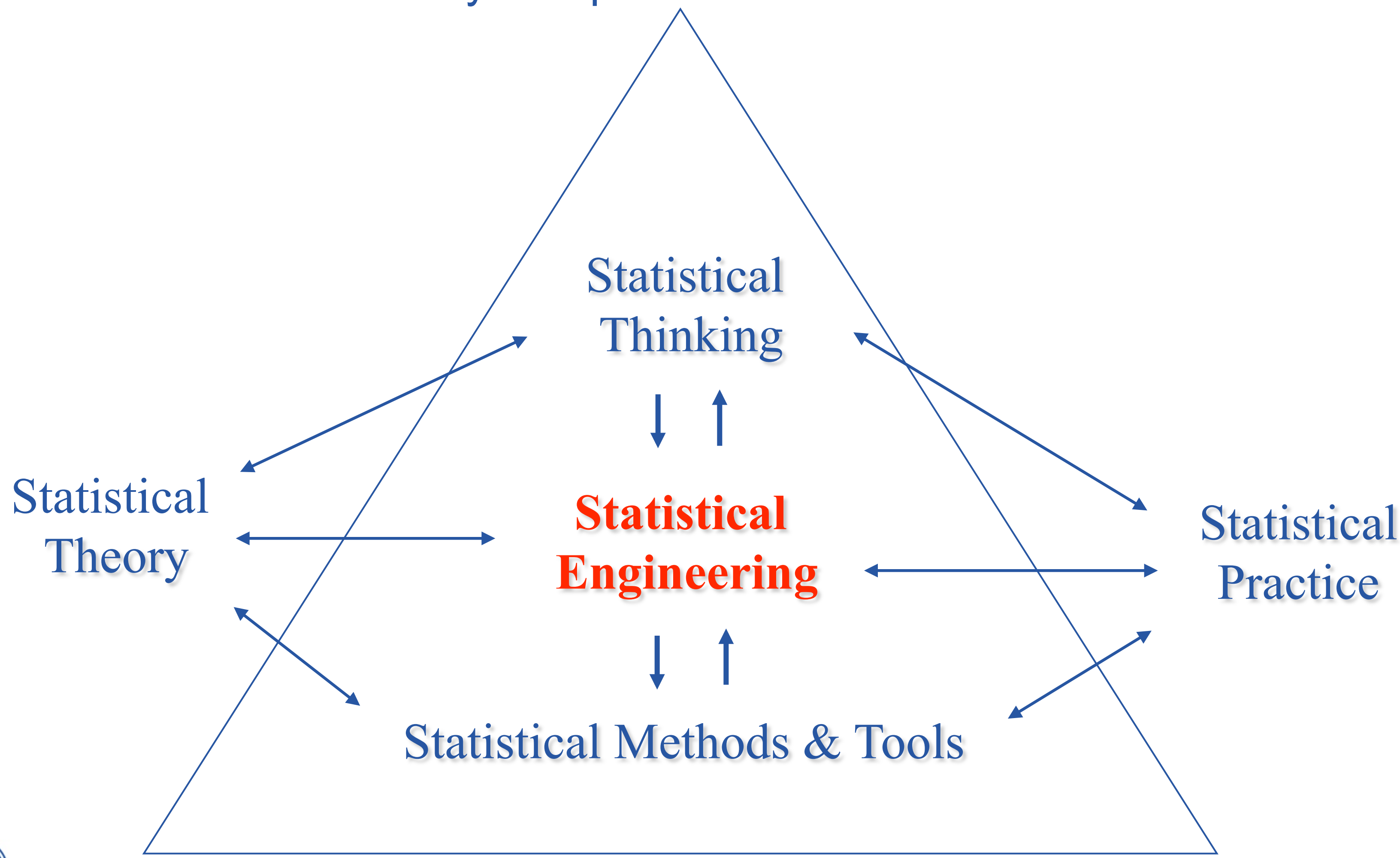
Potential Impact of Statistical Engineering

Statistical engineering is a potential bridge between statistical thinking and statistical tools and methods



Potential Impact of Statistical Engineering

We also believe that statistical engineering can help bridge the gap between statistical theory and practice



Large, Unstructured, Complex Problems

- The big payoff, mission critical problems that "do not correspond to an identifiable textbook chapter" (Meng 2009)
- Impact is broad – process performance, financial, customer, social and environmental
- Several departments, groups and functions are involved
- Problem has high degree of complexity involving both technical and non-technical challenges
- Multiple sources of data and information are used
- More than one statistical technique is required for solution
  - Typically non-statistical techniques are required
- Creative use of information technology is needed for success
- Long-term successes requires embedding solution into work processes typically through:
  - Use of custom software
  - Integration with other sciences and disciplines

Statistical engineering is needed for such problems – huge opportunity for the profession

My Message

To build on their successes and prosper in the 21st century, statisticians need to be more engaged in solving the big, complex, unsolved problems in their organizations

One key change needs to be a balanced focus on statistical engineering and statistical science

Statistical engineering can help:

- Ensure that statistical projects have high impact
- Provide a framework to attack large, complex, unstructured problems
- Integrate the principles of statistical thinking with the application of statistical methods and tools
- Enable statisticians to provide true leadership to their organizations,
  - Rather than focus on passive consulting services

I am anxious to hear what others have to say about statistical engineering this week.

References

Hoerl, R. W. and R. D. Snee (2002) *Statistical Thinking – Improving Business Performance*, Duxbury Press, Pacific Grove, CA.

Hoerl, R. W. and R. D. Snee (2009) "Post Financial Meltdown: What Do Services Industries Need from Us Now?", *Applied Stochastic Models in Business and Industry*, December 2009.

Hoerl, R. W. and R. D. Snee (2010a) "Moving the Statistics Profession Forward to the Next Level", *The American Statistician*, February 2010, 10-14.

Hoerl, R. W. and R.D. Snee (2010b) "Closing the Gap: Statistical Engineering can Bridge Statistical Thinking with Methods and Tools", *Quality Progress*, May 2010, 52-53.

Hoerl, R. W. and R. D. Snee (2010c) "Tried and True – Organizations put Statistical Engineering to the Test and See Real Results", *Quality Progress*, June 2010, 58-60.

Lindsay, B.G, Kettenring, J., and Siegmund, D.O. (2004) "A Report on the Future of Statistics (with Discussion)", Statistical Science, 19, 3, 387-413.

Meng, X. (2009) "Desired and Feared – What Do We Do Now and Over the Next 50 Years?", The American Statistician, 202-210.

Snee, R. D. and R. W. Hoerl (2003) Leading Six Sigma – A Step by Step Guide Based on Experience With General Electric and Other Six Sigma Companies, FT Prentice Hall, New York, NY,

Technometrics (2008) "Future of Industrial Statistics – A Panel Discussion". Technometrics Blog Link asq.org/discussionBoards/forum.ispa?forumID=77

Increasing Statistical Rigor in Operational Test and Evaluation

Dr. Michael Gilmore

Director, Operational Test & Evaluation

Keynote Address to the NASA Statistical Engineering Symposium

Outline

- DOT&E Background
- National Research Council Study
  - Statistics, Testing, and Defense Acquisition (1998)
- Current DOT&E Initiatives
- Increasing Statistical Rigor in Test & Evaluation
  - Design of Experiments
  - Reliability Growth

DOT&E Background

- DOT&E was created by Congress in 1983.
- Director is appointed by the President and confirmed by the Senate.
- Director’s reports, by statute, go directly to the Secretary of Defense and Congress
- Responsible for all operational test and evaluation, and live fire test and evaluation within DoD.
- Provides independent oversight and reporting.

Office of the Secretary of Defense

- Secretary of Defense
- Deputy Secretary of Defense
  - USD (Policy)
    - PDUSD (P)
  - USD (Comptroller)
  - USD (Personnel & Readiness)
  - USD (Intelligence)
  - **Director Operational Test & Evaluation**
  - USD (Acquisition, Technology & Logistics)
    - DUSD (Acquisition & Technology)
    - Director Defense Research & Engineering
    - DUSD (Logistics & Materiel Readiness)
    - DUSD (Acquisition & Technology)

The Director of Operational Test and Evaluation is the Principal Staff Assistant and advisor to the Secretary of Defense on all matters pertaining to operational test and evaluation within the US DoD.

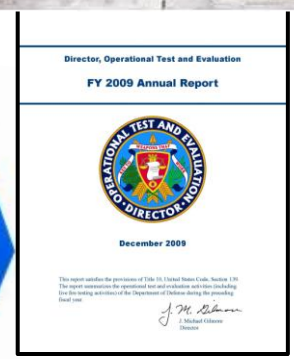
DOT&E Interactions

DOT&E Tools:

1. Test and Evaluation Master Plan approval
2. Test plan and Test Strategy approval
3. Beyond Low Rate Initial Production Reports
4. Early Fielding reports
5. Annual Report
6. Director's Memo, Testimony, Speeches
7. Close cooperation with Service Test
    Agencies

Guidance

BLRIPs, annual reports, testimony, etc.


Director, Operational Test and Evaluation

FY 2009 Annual Report

December 2009

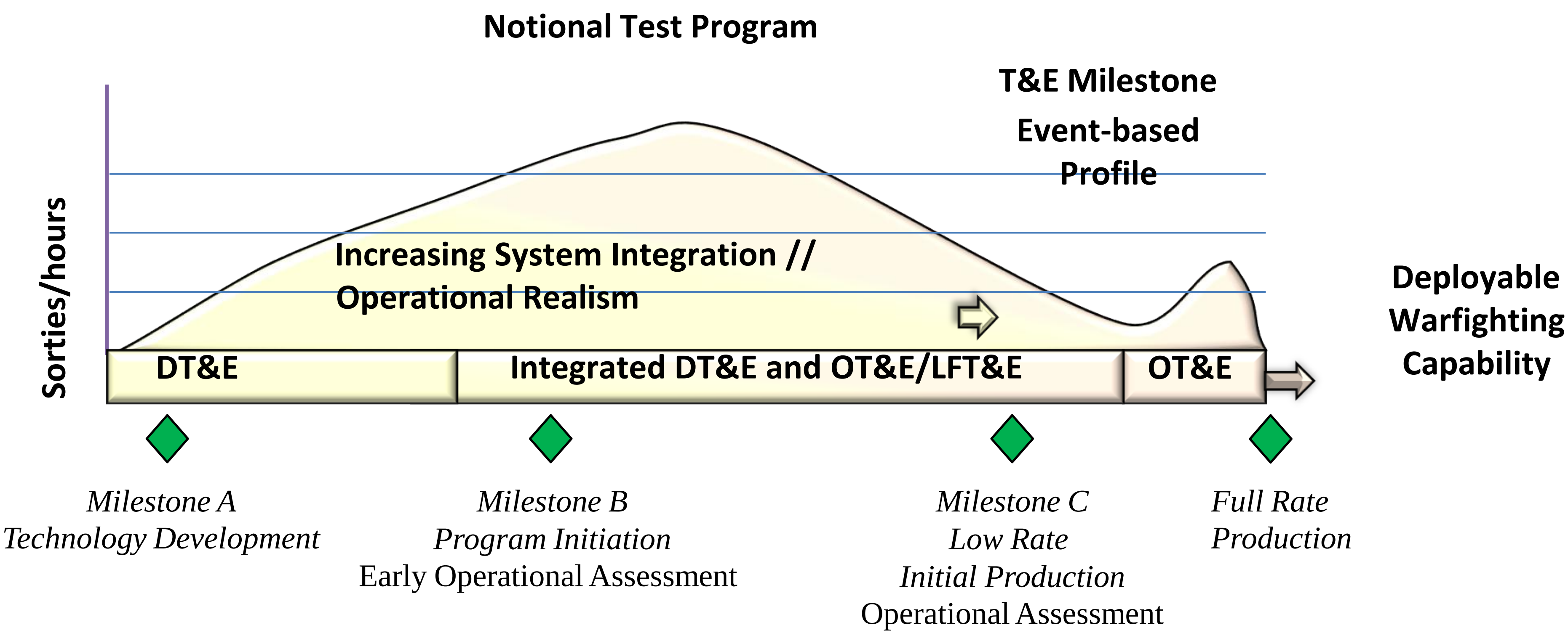
Guidance and consultation

DOT&E Responsibilities

- Prescribe Department of Defense policy for:
  - Operational Test & Evaluation (OT&E)
  - Live Fire Test & Evaluation (LFT&E)
- Provide guidance on all OT&E and LFT&E matters
- Monitor & review all OT&E and LFT&E
- Report annually to Congress on OT&E and LFT&E
- Member of Defense Acquisition Board
- Approve test plans for OT & LF oversight programs
- Report on programs, <u>before full-rate production decision</u> to the Secretary, OSD, Services, & four congressional committees:
    - Adequacy operational and live fire testing
    - Operational Effectiveness
    - Operational Suitability
    - Survivability and Lethality

Acquisition Timeline

- Operational Testing supports full rate production decision
- Report on programs, before full-rate production decision:
  - Test adequacy, Operational Effectiveness, Suitability, Survivability and Lethality



Why is a Program Placed on DOT&E OT Oversight?

- Statutory requirement for Major Defense Acquisition Programs (MDAPs)
  - Designated by the Secretary of Defense, or
  - Estimated by the Secretary of Defense to require an eventual total expenditure for research, development, test, and evaluation of more than $300,000,000 (based on fiscal year 1990 constant dollars), or
  - Estimated to have an eventual total expenditure for procurement of more than $1,800,000,000 (based on fiscal year 1990 constant dollars).
- Congress or OSD has expressed a high level of interest
- Congress has directed DOT&E report as condition for production or progress
- Program requires joint or multi-Service testing
- Program has a close relationship or is key to a major program
- Militarily significant change to system

National Research Council Study (1998)

- Panel on Statistical Methods for Testing and Evaluating Defense Systems
  - Examine statistical techniques & make recommendations
- Statistics, Testing, and Defense Acquisition: New Approaches and Methodological Improvements
- Select conclusions & recommendations
  - Conclusion 3.1: Major advances can be realized by applying selected industrial principles and practices in restructuring the paradigm for operational testing…
  - Conclusion 4.1: The current practice of statistics in defense testing design and evaluation does not take full advantage of the benefits available from the use of state-of-the-art statistical methodology.
  - Recommendation 4.2: All estimates of the performance of a system from operational test should be accompanied by statements of uncertainty through use of confidence intervals…
  - Recommendation 4.4: The service test agencies should examine the applicability of state-of-the-art experimental design techniques and principles and, as appropriate, make greater use of them in the design of operational tests.
  - Recommendation 7.4: Operational test agencies should promote more critical attention to the specification of statistical models of equipment reliability, availability, and maintainability and to supporting the underlying assumptions…
- The majority of the recommendations have still not been implemented 13 years later.

DOT&E Initiatives

Integrated Testing
Developmental
Operational
Live Fire

Testers Engage Early
Requirements are:
Unambiguous
Testable
Relevant to mission accomplishment
Technically Realistic

Improve Suitability
Reliability Growth

Field New Capabilities Rapidly
Accelerated Testing

DOT&E Initiatives & Increasing Test Rigor

- Integrated Testing
  - "… DOE provides the scientific and statistical methods needed to rigorously plan and execute tests and evaluate their results. … The DT&E and OT&E offices are working with the OTAs and Developmental Test Centers to **apply DOE across** the **whole development and operational test cycle** for a program.
  - "DOE should allow DOT&E to make statements of the confidence levels we have in the results of the testing. Whenever possible, our evaluation of performance must include a rigorous **assessment of** the **confidence level of the test**, the **power of the test** and some measure of how well the **test spans the operational envelope** of the system."

Apply DOE Across Entire Acquisition Development Cycle

Design of Experiments (DOE)

- Test planning is a science
- DOT&E must evaluate test plan adequacy
  - TEMP
  - IOT&E
- Statistics equips us to determine:
  - Breadth of coverage
  - Power
  - Confidence
- Design of Experiments is a formal scientifically based method for constructing test plans.
  - There are many tools within the DOE toolbox.
  - Key idea behind DOE: strategically manipulate factors & levels (inputs) to influence the output.

Measures of test plan adequacy!

DOE is a scientific tool for developing robust test plans!

DOT&E Retrospective Case Studies

- Motivation: get beyond general concepts, look at real world examples
- Goal: provide baseline and highlight areas for improvement
- Conducted an analysis of select Beyond Low Rate Initial Production (BLRIPs) from last two years
  - Noted a structured approach to testing that captures many aspects of these concepts
  - The analysis also identified areas of potential improvement

Case Studies – BLRIP Reviews

- USS Virginia
- Guided Multiple Launch Rocket System (GMLRS) Unitary
- MH-60R and MH-60S
- **Stryker Mobile Gun System (MGS)**
- **Joint Chemical Agent Detector (JCAD)**
- DoN Large Aircraft Infrared Countermeasures (LAIRCM)
- EA-18G

Stryker Mobile Gun System (MGS)
Coverage of Operational Envelope

4 Factors: Mission Type, Terrain Type, Threat Level & Illumination

	Mission	Attack				Defend				Stability and Support				
Illum	OPFOR	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	
Day	Low	1	1											2
Day	Med	1				1								2
Day	High	1				1	3							5
Night	Low										2			2
Night	Med	2												2
Night	High		2			1								3
		5	3			3	3				2			16

Weather: as it occurred; not controlled

Key

- Instrumented data collected during controlled IOT at Ft. Hood; number of mission replications indicated in cell

- Limited use data collected during Mission Rehearsal Exercise at Ft. Lewis; no instrumentation or control over factors

- Limited use (anecdotal) data collected in theater during unit deployment to OIF, mostly on tactics and employment techniques

- IOT test design builds on evidence from previous events
  - Mission Rehearsal Exercise prior to unit deployment (basis for Section 231 report)
  - Field data from unit deployment
- IOT scoped to focus on voids in medium and high threat levels

Early deployment changed original DOE plan

Stryker Mobile Gun System (MGS)

Lesson Learned:
"DOE" identified gaps in coverage,
partially filled from other sources

4 Factors: Missio[n] ... Level & Illumination

	Mission	Attack				Defend				Stability and Support				
Illum	OPFOR	Urban	Mixed	Forest	Des[ert]	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	
Day	Low	1	1											2
Day	Med	1				1								2
Day	High	1				1	3							5
Night	Low										2			2
Night	Med	2												2
Night	High		2			1								3
		5	3			3	3				2			16

Weather: as it occurred; not controlled

Key

- Instrumented data collected during controlled IOT at Ft. Hood; number of mission replications indicated in cell
- Limited use data collected during Mission Rehearsal Exercise at Ft. Lewis; no instrumentation or control over factors
- Limited use (anecdotal) data collected in theater during unit deployment to OIF, mostly on tactics and employment techniques

- IOT test design builds on evidence from previous events
  - Mission Rehearsal Exercise prior to unit deployment (basis for Section 231 report)
  - Field data from unit deployment
- IOT scoped to focus on voids in medium and high threat levels

Early deployment changed original DOE plan

Impact of Experimental Design

- Case Study: Mobile Gun System Design Comparison

	Executed Cases in IOT&E	DOE I - Factorial Design	DOE II – Optimal Design (large)	DOE III – Optimal Design (small)
Factors & Levels	4 factors: Mission Type (3), Terrain Type (4), Treat Level (3), Illumination (2)			
Total Tests: N	16	72	36	16
Confidence: (1-α)	Set to the same level across all 4 designs: Confidence = 95%			
1σ - Power (1-β)	8.1% - 28.0%	68.1% - 98.7%	35.0% - 81.5%	12.3% - 39.9%

- The case study suggests that 16 runs is far from adequate to span the operational battle space with high power and confidence.
- The DOE optimal design is a more powerful allocation of the 16 tests than the case based design.
- DOE allows us to understand what we are giving up!
  - In the case of MGS, the system was deployed early which altered the original test plan.

MGS Case Design vs. D-Optimal Design

Case Based Design Executed in IOT&E

	Mission	Attack				Defend				Stability and Support				
Illum	OPFOR	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	
Day	Low	1	1											2
Day	Med	1				1								2
Day	High	1				1	3							5
Night	Low										2			2
Night	Med	2												2
Night	High		2			1								3
		5	3			3	3				2			16

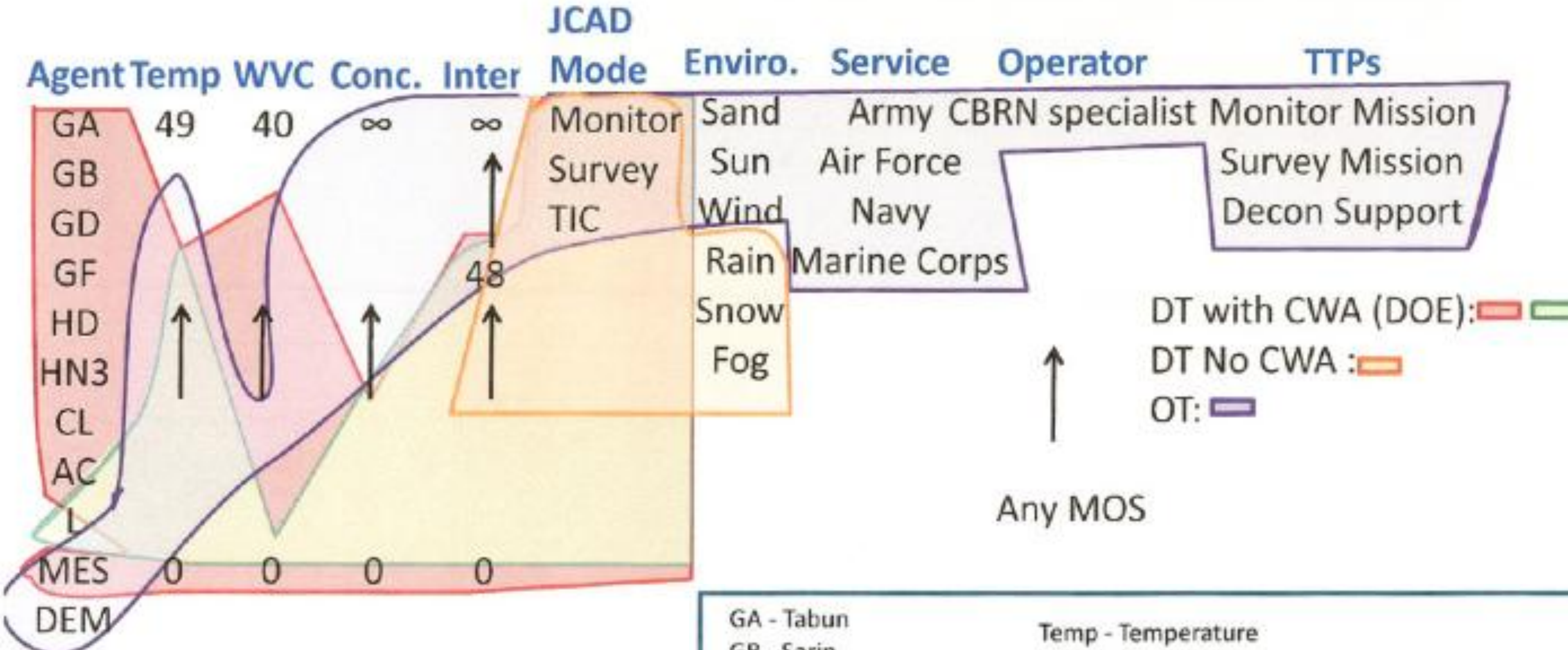
Statistical D-Optimal Design

	Mission	Attack				Defend				Stability and Support				
Illum	OPFOR	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert	
Day	Low					1		1					1	3
Day	Med		1		1									2
Day	High	1						1			1			3
Night	Low		1	1										2
Night	Med					1	1					1		3
Night	High				1				1	1				3
		1	2	1	2	2	1	2	1	1	1	1	1	16

Joint Chemical Agent Detector (JCAD)

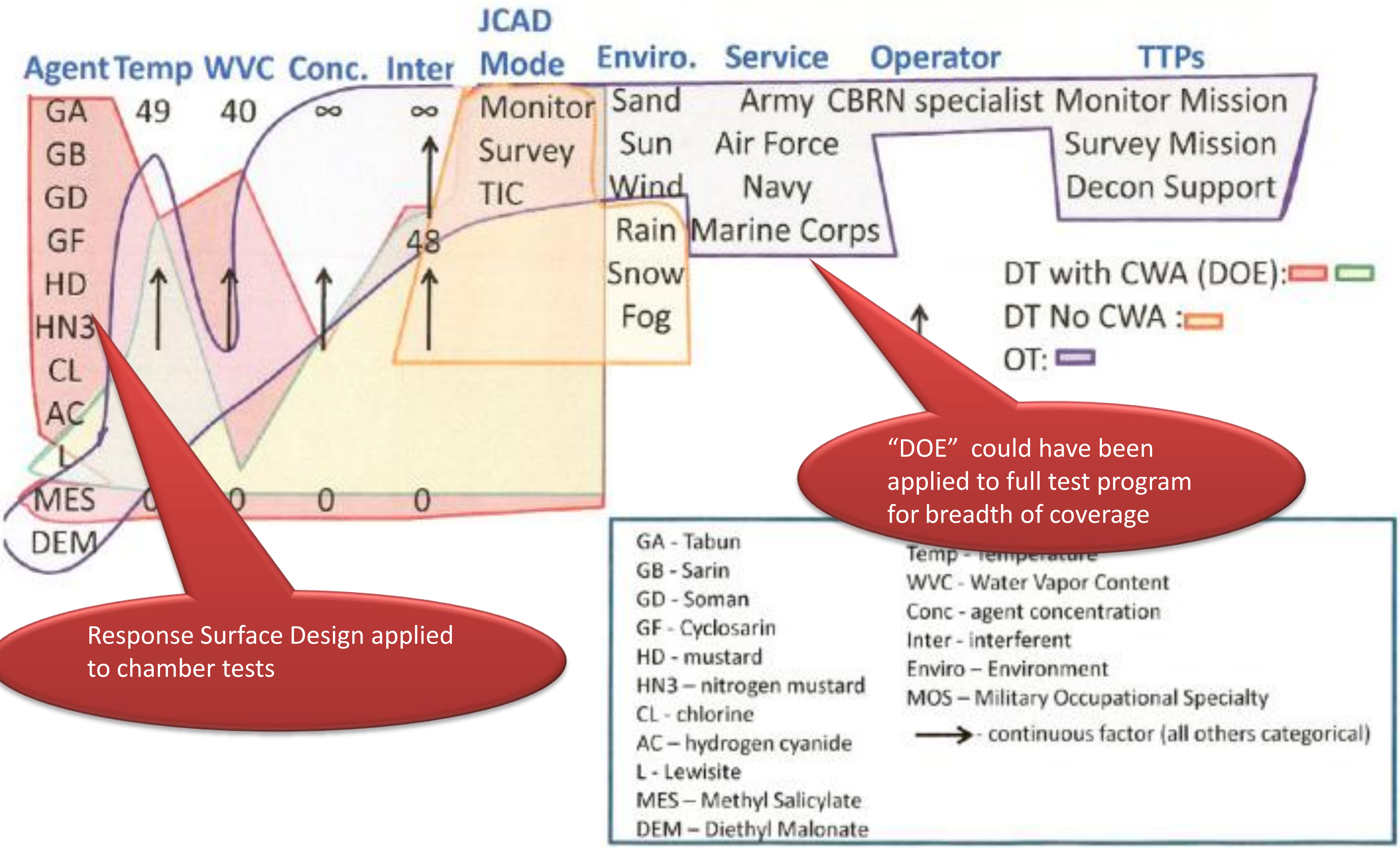
- What is the operational envelope? (factors and levels)
  - Agent (9 agents and 2 simulants)
  - Temperature, water vapor concentration, agent concentration, interferent (continuous)
  - Environment (sand, sun, wind, rain, snow, fog)
  - Service (Army, Air Force, Navy, Marine Corps)
  - JCAD Mode (Monitor, Survey, TIC)
  - Operator (Any MOS to CBRN Specialist)
  - TTP (Monitor Mission, Survey Mission, Decon Support)

Joint Chemical Agent Detector (JCAD) Coverage of Operational Envelope



- GA - Tabun
- GB - Sarin
- GD - Soman
- GF - Cyclosarin
- HD - mustard
- HN3 – nitrogen mustard
- CL - chlorine
- AC – hydrogen cyanide
- L - Lewisite
- MES – Methyl Salicylate
- DEM – Diethyl Malonate
- Temp - Temperature
- WVC - Water Vapor Content
- Conc - agent concentration
- Inter - interferent
- Enviro – Environment
- MOS – Military Occupational Specialty
- → - continuous factor (all others categorical)

Joint Chemical Agent Detector (JCAD) Coverage of Operational Envelope



Joint Chemical Agent Detector (JCAD)
Power of Test

- Power Analysis for JCAD Chamber Test
  - DT Testing
  - Statistical Response Surface Design (I-Optimal)
  - High power test plan

Factor	S:N* = 0.5	S:N = 1.0	S:N = 2.0
Temperature	32.0%	84.7%	99.9%
Water Vapor Content (WVC)	42.1%	94.1%	99.9%
Concentration	46.5%	96.3%	99.9%

*S:N – signal-to-noise ratio, goal detectable difference as a ratio to the design standard deviation

Lessons Learned – Case Studies

- There is no “one size fits all” approach for designing experiments.
- Review of past BLRIPS illustrated the need for a more rigorous approach to DT/OT incorporating statistical experimental design.
- Defining measurable, testable, important response variables is key to a good design.
- T&E already thinks about factors and levels (input variables); however we can improve on a scientific approach to varying factors & levels.
- DOE can be applied in both DT and OT. However, responses and factors may differ between the two.
- Confidence and power add important information to our assessment of the test adequacy and test results.

The Strength of Designed Experiments

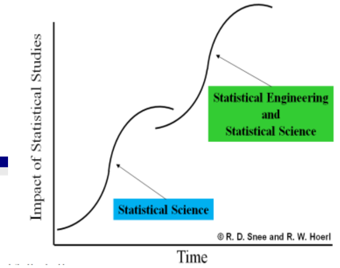
- We face the following constraints in testing:
  - Limited Test Resources
    - Range Size
    - Forces (Manpower & Materiel)
  - Limited Test Time
    - Fielding & Production Schedules
    - Range Availability
  - Limited Test Articles
    - Unit Cost
    - Production Time
- Design of experiments allows us to understand the tradeoffs these constraints impose.
- Design of experiments can provide a statistically optimum allocation of our assets under given constraints.

DOT&E Guidance on Design of Experiments

- DOT&E Memorandum, October 19, 2010
- "Design of Experiments is a structured process to identify the metrics, factors, and levels that most directly affect operational effectiveness and suitability…"
- Elements of experimental design for TEMPs and Test Plans approval:
  - The goal of the experiment.
  - Quantitative mission-oriented response variables
    - effectiveness and suitability
  - Factors [and levels] that affect response variables
  - A method for strategically varying factors
    - across both developmental and operational testing
  - Statistical measures of merit
    - power and confidence on the relevant response variables for which it makes sense.
    - Important to understand "how much testing is enough?"

Challenges to Integrated Testing using DOE

- Current policy documents do not address the use of statistics during T&E planning or analysis.
  - No Policy → No funding
- Program Managers decide how much and where DT is done
  - Makes planning sequential experiments challenging.
- Developmental testing has not been sufficient or adequate
  - OT&E results indicate a Department-wide problem
  - Seeing more weapons systems not ready for IOT&E and combat.
- Congress recently created a Director of Developmental Test and Evaluation (DDT&E)
  - DDT&E has not yet provided guidance on the application of statistical methods in DT



Leadership – Essential for Developing the Discipline of Statistical Engineering

Ronald D. Snee

Snee Associates, LLC

With Significant Contributions by

Roger Hoerl

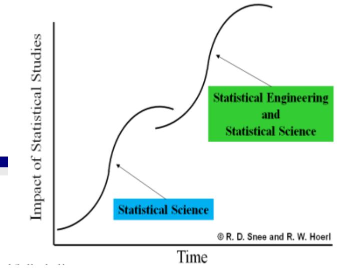
GE Global Research

NASA Symposium on Statistical Engineering

Williamsburg, VA

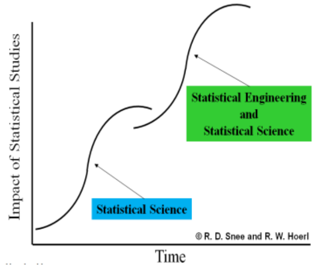
May 3-5, 2011

Agenda



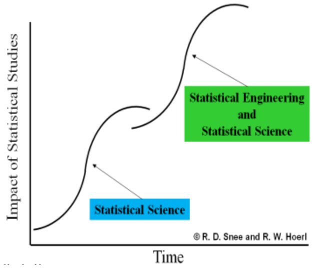
- Today's Realities
- Increasing the Impact of our Work
- Dealing with Large, Unstructured, Complex Problems
- Why Leadership is Needed
  - What Does Management Expect of Statisticians and other professionals?
- What Do Leaders Do?
- Developing Leadership Skills
- Conclusion

Today's Realities



- Global Competition, fueled by information technology, is forcing changes in all aspects of our society
  - Business
  - Government
  - Education
  - Health Care
- Customers are demanding more
- We have to change how we work and manage
  - All aspects of our organizations
  - All processes we use to do our work

Implications for Use of Statistical Thinking and Methods

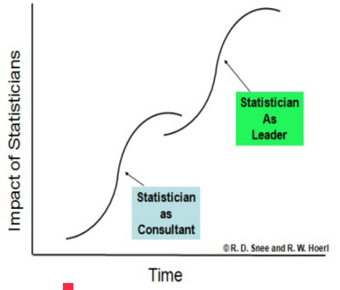


- Internet enables communication and data access available to anyone in the world at any time.
- Anyone with a PC can download and use statistical software
- Cost conscious organizations are asking why employees can't analyze their own data
- Narrow technical tasks, such as intensive data analysis, are easy to transfer to employees in low cost countries*
- The "market" for the old role of a consultant who performs data analysis is rapidly evaporating.

Organizational Leaders Are Looking for People Who Can Deliver High Impact Results

*We do not suggest that this controversial phenomenon is either good or bad; only that it is occurring.

Expanding Role of Statisticians



Consultant/Expert

- Consult on other people's projects
- Perform routine analyses if needed
- Teach statistical tools
- Work with technical people

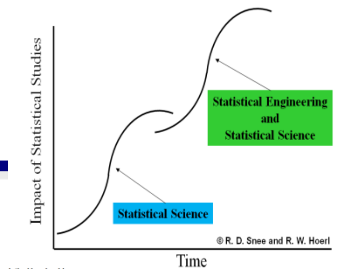
- Narrow expertise and accountability
- "Benign neglect"

Collaborator/Leader

- Lead or collaborate on our own projects – project ownership
- Focus on significant, complex problems
- Design training systems
- Work with managers and technical people
- Broad expertise and accountability
- "In the firing line"

Computer Scientists Provide an Example of Such a Role

Statistical Engineering Is A Team Sport

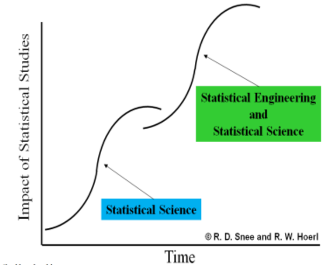


Management and Organizational Leaders

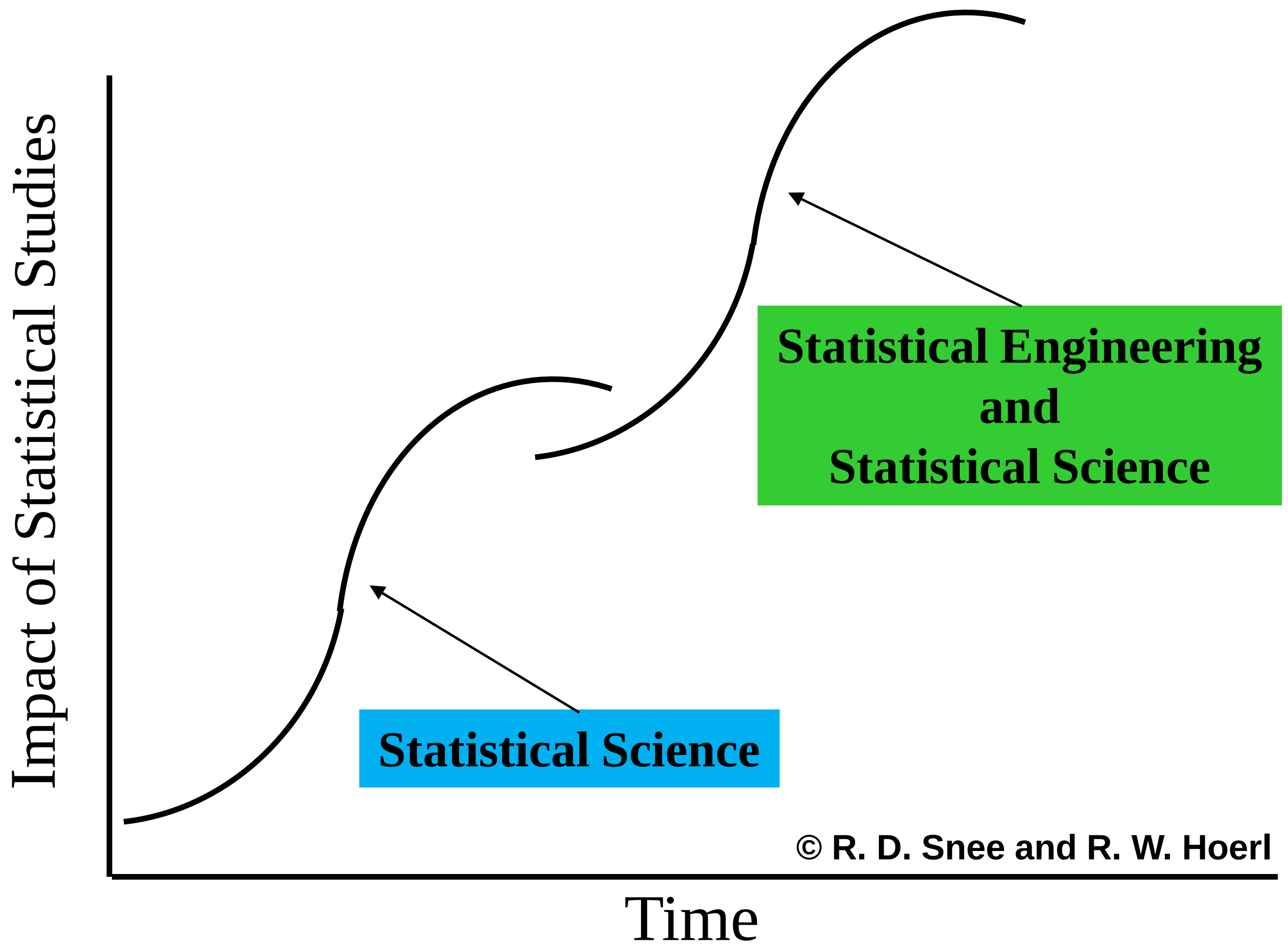
Statisticians

Scientists, Engineers and Other Professionals

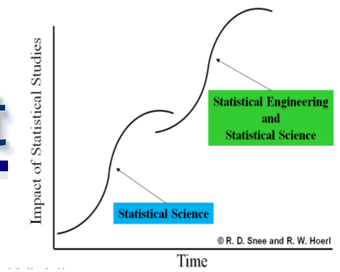
Statistical Engineering Increases Impact



Balanced approach involving both statistical science and statistical engineering will enable us to have a much broader impact on society

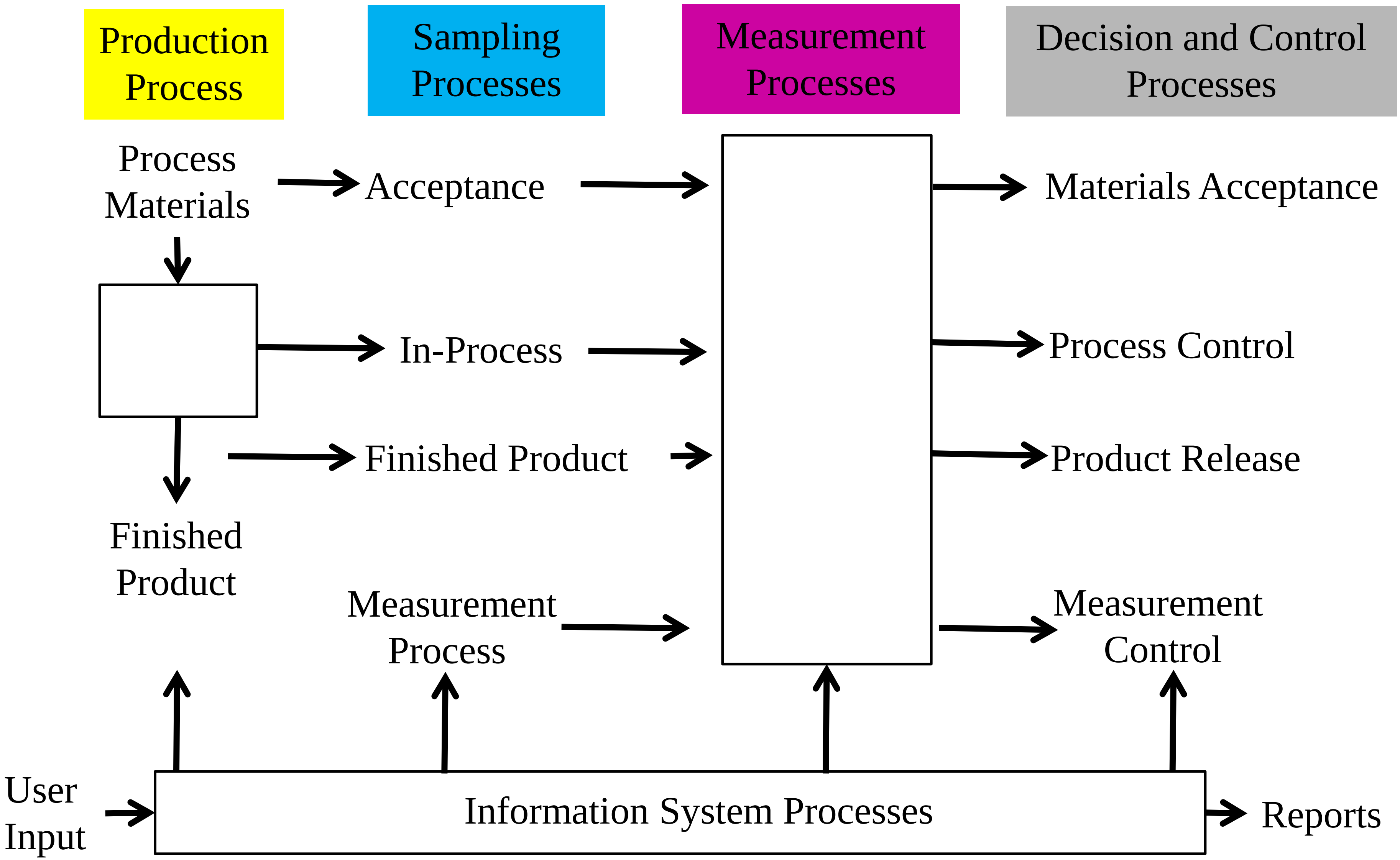


Case Study – Product Quality Management

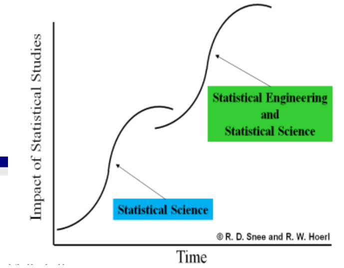


- Background
  - DuPont Dacron polyester products experiencing quality problems in the marketplace
  - Products produced at 8 plants globally
  - Key players – Manufacturing and Marketing – not working together
  - Many quality and statistical tools are available to solve this problem
- Questions
  - Is this an opportunity with major impact?
  - Is this a large, unstructured, complex problem?
  - What is needed to be successful?

PQM System


Production Process
Sampling Processes
Measurement Processes
Decision and Control Processes
Process Materials
Acceptance
Materials Acceptance
In-Process
Process Control
Finished Product
Product Release
Finished Product
Measurement Process
Measurement Control
User Input
Information System Processes
Reports

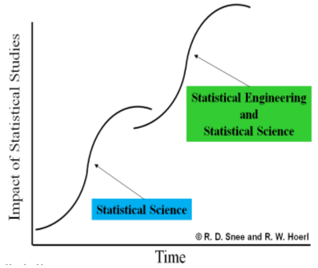
Case Study - PQM System – Statistical Techniques



- Product Release
  - Sampling Schemes
- Process Control
  - Cumulative Sum
- Measurement Variation
  - Shewhart and CUSUM Control charts
  - Inter-laboratory studies
- Process Calibration and Adjustment
  - Response Surface Methods
- Process Variance Components - Estimation and maintenance

Statistical Techniques Were Integrated, Linked and Sequenced to Produce the Product Quality Management System

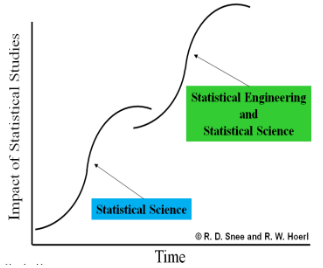
DuPont PQM: Statistically-Based Product Quality Management System



- Product Quality Management (PQM)
  - Framework for managing the quality of a product or service.
  - Operational system the enables Marketing, R&D, Production and support personnel to work together to meet increasingly stringent customer requirements
- "Within two years product quality had improved to the point of commanding a marketplace advantage and more than $30 million had been gained in operating cost improvements. The statistically based Product Quality Management system developed for "Dacron" was expanded to other products with further contributions in earnings."

Richard E. Heckert
Chairman and CEO, DuPont Company
ASA Annual Meeting 1986

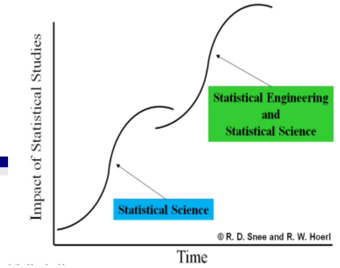
Large, Unstructured, Complex Problems



- The big payoff, mission critical problems
- Impact is broad
  - Process performance, financial, customer, social and environmental
- Several departments, groups and functions and disciplines are involved
- Problem has high degree of complexity
  - Technical and non-technical challenges are involved
- Multiple sources of data and information are used
- Mix of statistical and non-statistical techniques are required for solution
- Creative use of information technology is needed for success
- Long-term successes requires embedding solution into work processes

Statistical Engineering is Needed for Such Problems
Huge Opportunity for All Involved

Case Study
Product Quality Management System

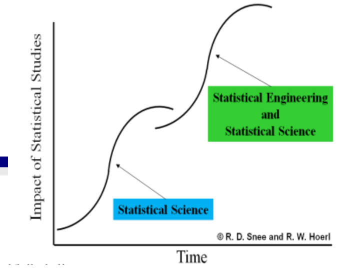


Role of Statisticians in PQM at DuPont

- Leaders in developing their organization's strategy for quality management
- Leaders in developing technology systems for quality management
- Participants in the business planning process
- Participants in problem solving activities
- Leaders in initiating and implementing quality management training systems at all levels

DuPont PQM Manual 1988

What Do Executives Expect?



Executives Interviewed (Vining, Bowen and Parr 2001)

- CEO, Major Technology Co., #1
- CEO, Major Service Company
- Senior VP, Major Manufacturing Co.
- VP, R&D, Major Food Company
- VP, Major Technology Co., #2
- VP and GM, Major Manufacturing Co.

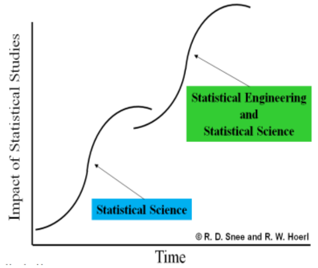
Executives Expectations

- Persons Who Can Lead Projects and Get Results

"Results Buy Freedom"

Arnie Eckleman,
Senior Vice President
Verizon Communications

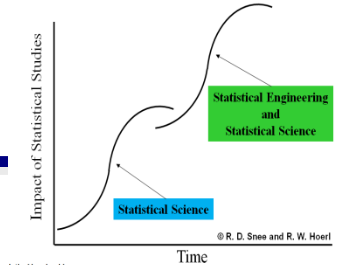
Developing Leadership Skills



Some Questions You May Have

- What is Leadership?
- What Do Leaders Do?
- Key Leadership Skills?
- How do I develop leadership skills?

Leaders Help Us Make the Needed Changes?



Leaders help a group of people move from

One paradigm to another

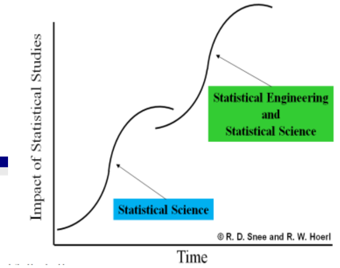
"Leadership: the art of getting someone else to do something you want done because he wants to do it."

Dwight D. Eisenhower

"Leaders have followers"

Bill Gore, Founder
W. L. Gore and Associates

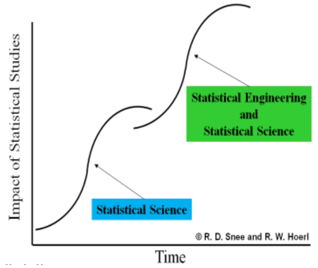
We have Many Kinds of Leaders



- Political
- Military
- Business
- Academic
- Religious
- Sports
- Statistical Leaders
  - Technical
  - Managerial
- And many more

……………

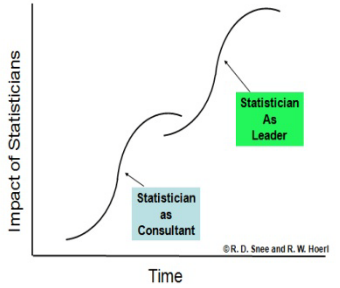
Myths of Leadership



- Leadership is a rare skill
- Leaders are born not made
- Leaders are charismatic
- Leadership exists only at the top of an organization
- The leader controls, directs, prods, manipulates

Bennis and Nanus 1985

Change Requires Both Leading and Managing



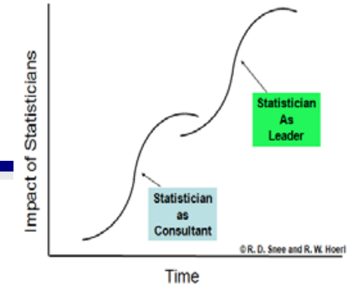
Leading

- Moving Between Paradigms
- Doing Right Things
- Creating Improvements
- Leading & Developing People

Managing

- Working Within a Paradigm
- Doing Things Right
- Holding the Gains
- Managing Processes

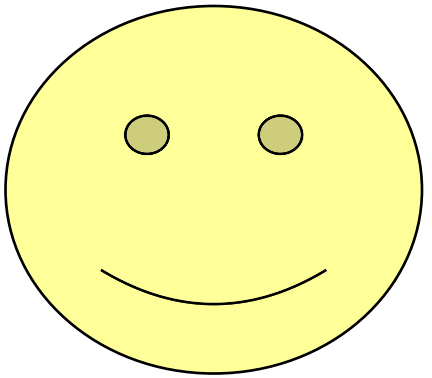
We Need Both Leading and Managing



Lead People

- Direction
- Communication
- Resources
- Coach
- Reinforce

Manage Processes



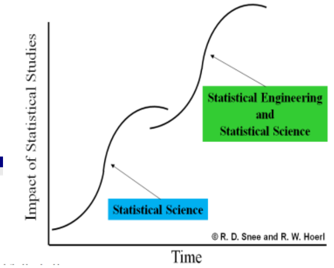
Happy Customers

Feedback

- What's working?
- Need to do differently?

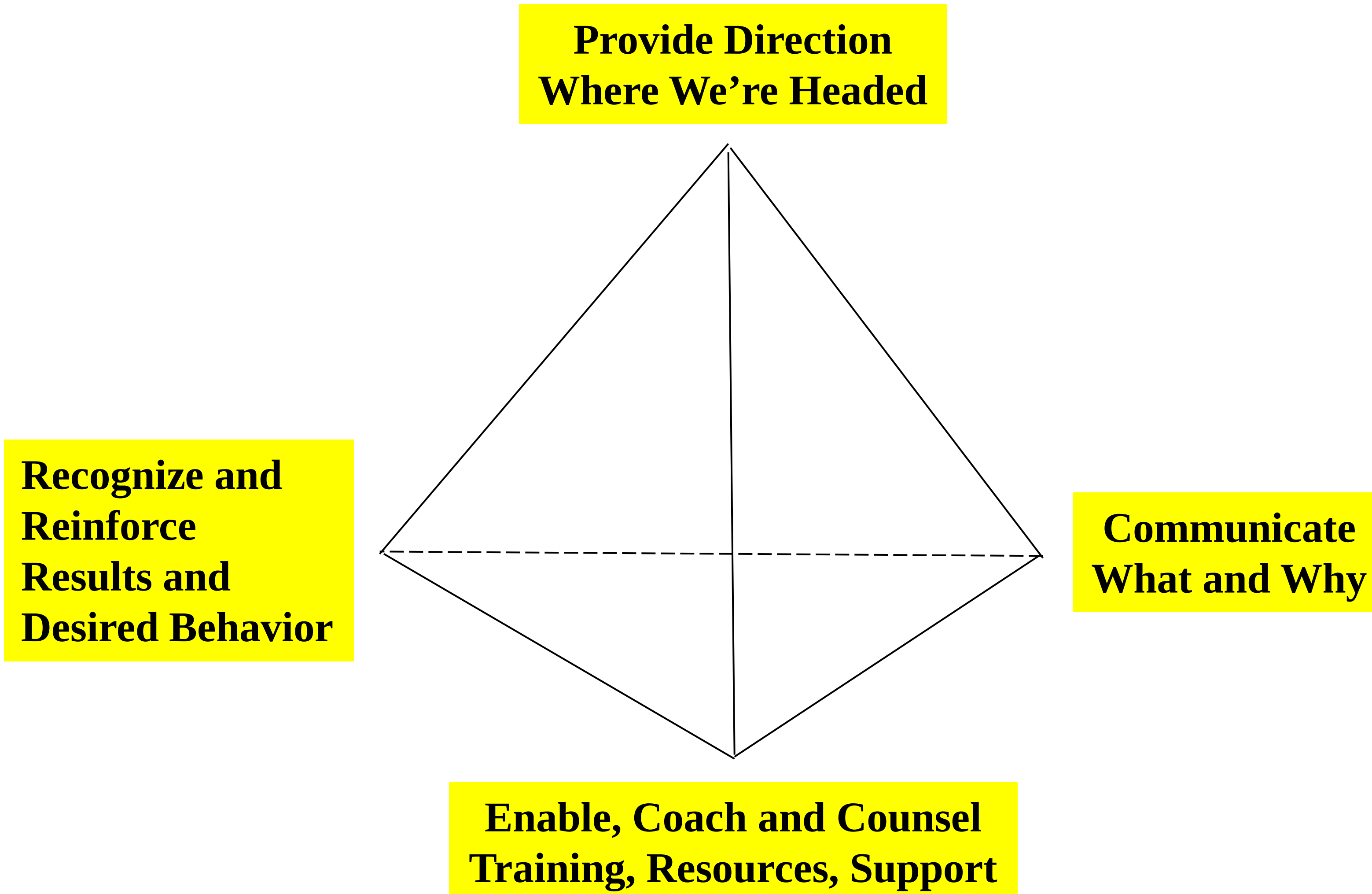
Healthy processes serving happy customers

Time Spent on Doing and Improving Work

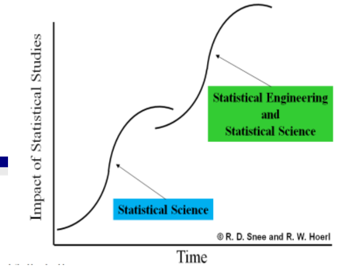


Role	Leading "Improving" Work	Managing "Doing" Work
Executives	90	10
Managers	70	30
Others	30	70

So What Do Leaders Do?

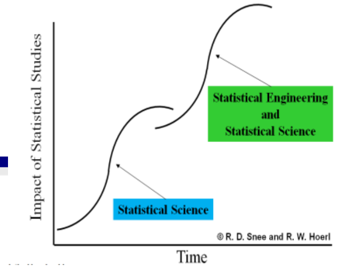


Providing Direction - Showing the Way



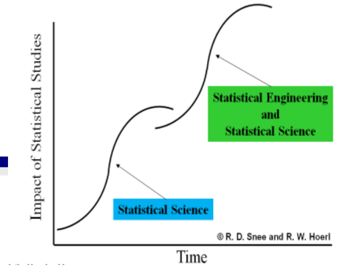
- Vision - What Success Looks Like
- Objectives - How we will win
- Goals - How much, by when
- Strategies - What we will focus on
- Initiatives - Specific projects we will undertake

Kotter's Eight Stages of Successful Change



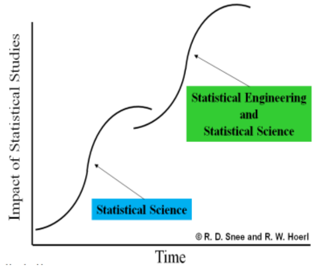
- Establish a sense of urgency
- Create a guiding coalition
- Develop a vision and strategy
- Communicate the change vision
- Empower employees for broad based action
- Generate short-term wins
- Consolidate gains and produce more change
- Anchor new approaches in the culture

Communicate The Direction
Provide Understanding and Hope



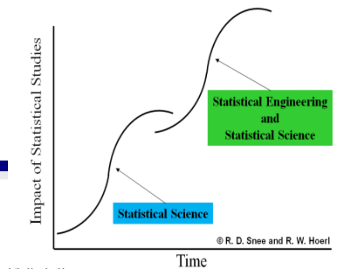
- The direction we are pursuing
- What benefits we expect to get
- Progress - Results achieved to date
- Communication should be clear, concise and continuous
- Variety of media should be used
  - People take in and process information in different ways

Enable - Set Up People for Success



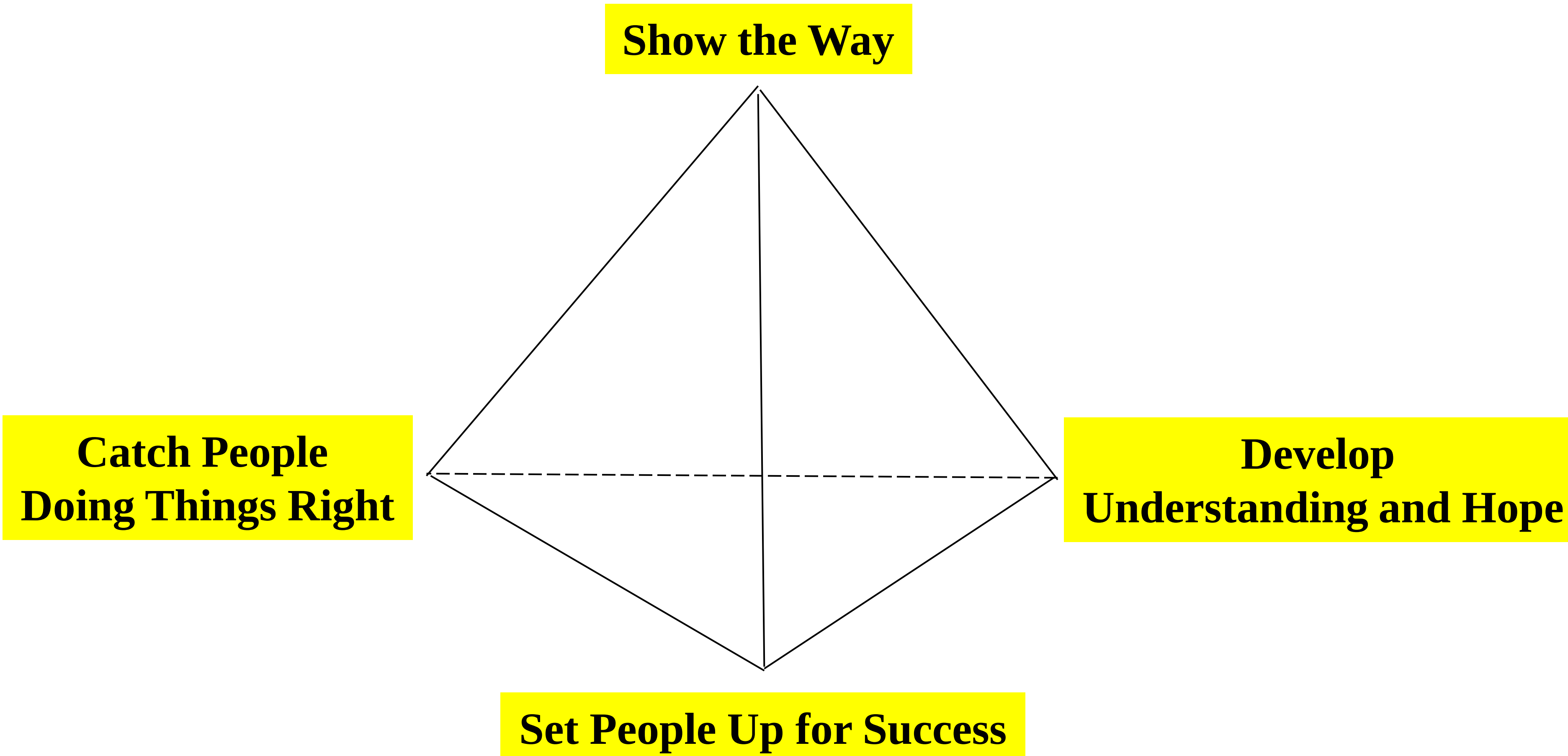
- Provide resources - people, time, $$$
- Provide training - build needed skills
- Provide methods to accomplish assigned tasks
- Remove barriers
- Coach and Counsel

Recognize and Reinforce –
Catch People Doing Things Right

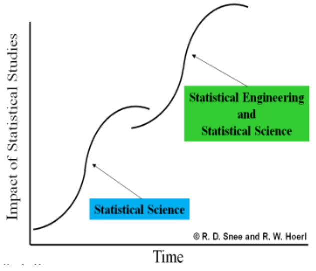


- Recognize accomplishments and results
  - Psychological rewards
  - Financial rewards
- Reinforce desired behavior
  - Catch people doing things right
- People want and need feedback
  - "How am I doing?" , Ed Koch, Mayor, New York
- Feedback needed for improvement
- Key tool - Management reviews

Leaders Lead People – Leaders ….

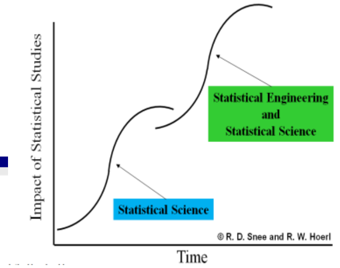


Critical Leadership Skills



- Organizational Acumen
  - Understand how the organization works
- Process and Systems Thinking
- Strategic Planning and Deployment
- Stakeholder Building
- Communication – Clear, concise and continuous
- Reviewing and Coaching
- Structured Improvement Methods (DMAIC)
- Learn to Deal With Teams and Group Dynamics
- Meeting Design and Facilitation
- Project Planning and Management
- Understanding Human Behavior

Ways to Develop Leadership Skills

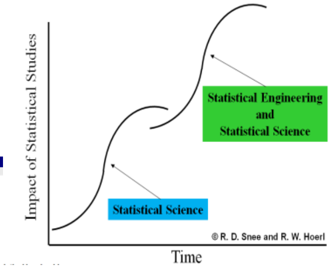


- Read books and articles
- Attend courses
- Discuss the subject with colleagues
- Practice, Practice, Practice, ………..

"Becoming a leader is like learning to play the violin in public"

Anonymous

My Message



- Place an increased focus on enhancing your leadership skills
- Be on the lookout for examples - good and bad - of leadership that you can use as a model
- Personal Change Is Required
  - Those Who Fail to Respond to Their Changing World Will Have Less Influence in It
- Best way to learn to lead is to do it

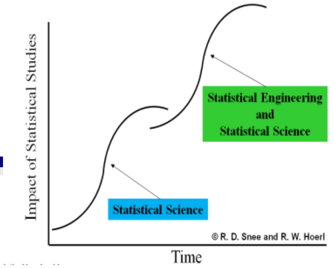
  - Be on the lookout for *your opportunity*!

Can We Be Successful?

"Never doubt that a small group of committed people can change the World. Indeed, it is the only thing that ever has"

Margaret Mead

Books on Leadership



Bennis, W and Nanus, B. (1985) Leaders - The Strategies for Taking Charge, Harper-Row, NY. Bennis has several other books on leadership.

Cohen, A. R and D. L. Bradford (2005) Influence Without Authority, Second Edition, John Wiley and Sons, New York, NY, 2005.

Covey, S. R. (1989) The Seven Habits of Highly Effective People", Fireside, Simon and Schuster, New York, NY

Hayword, F. (1997) Churchill on Leadership, Prima Publishing, Rocklin, CA

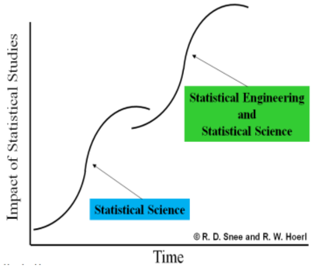
Kotter, J. (1997) Leading Change, Harvard Business School Press, Cambridge, MA. Also see Harvard Business Review, March-April 1995

Kouzes, J. M. and Posner, B. Z. (1995) The Leadership Challenge, Jossey-Bass, San Francisco

Phillips, D. T. (1992) Lincoln on Leadership, Warner Books, New York, NY

Snee, R. D. and R. W. Hoerl (2003) Leading Six Sigma – A Step-by-Step Guide Based on Experience With GE and Other Six Sigma Companies, FT Prentice Hall, Upper Saddle River, NJ

References on Leadership Skills



Bens, I (1999) Facilitation at a Glance, GOAL/QPC, Methuen, MA

Brassard, M. (1989) The Memory Jogger Plus, Goal/QPC, Methuen, MA

Dalton, G. W. and Thompson, P. H. (1986) Strategies for Career Management, Scott Foresman, New York, NY

Dolye, M. and Straus, D. (1982) Making Meetings Work, Jove Books, New York, NY

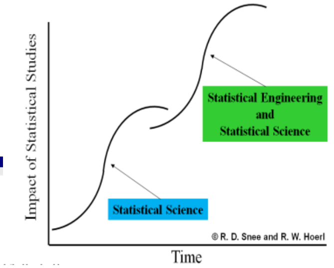
GOAL/QPC (1997) Project Management Memory Jogger, GOAL/QPC, Methuen, MA

Hoerl, R. W. and Snee, R.D. (2002) Statistical Thinking – Improving Business Performance, Duxbury Press, Pacific Grove, CA

Scholtes, P. R., Joiner, B. L. and Streibel, B. J. (1996) The Team Handbook, Oriel, Madison, WI

Snee, R. D. (1998) “Non-Statistical Skills that Can Help Statisticians Become More Effective”, Total Quality Management Journal, Vol. 9. No. 8, 711-722

Articles on Statistical Leadership



Anthony, J. and R. D. Snee (2010) “Leading Role – Identifying the Skills Master Black Belts and Black Belts need to be Effective Leaders”, Six Sigma Forum Magazine, May 2010, 6-12.

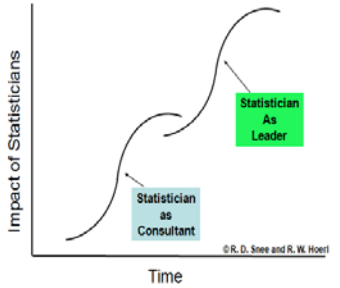
Hoerl, R. W. and R. D. Snee (2010) “Moving the Statistics Profession Forward to the Next Level”, The American Statistician, February 2010, 10-14.

Snee, R. D. and R. W. Hoerl (2004) “Statistical Leadership – as Traditional Workplace Roles Change, Learn to Transition from Consultant to Leader”, Quality Progress, October 2004, 83-85.

Snee, R. D. (2005) “Leading Business Improvement: A New Role for Statisticians and Quality Professionals”, Quality and Reliability Engineering International, Volume 21, 2005, 235-242.

Snee, R. D. (2007) “Turning Shewhart’s Challenge into Opportunity – Statisticians Must Step Forward and Lead Management to Become More Statistically –Minded”, Quality Progress, May 2007, 70-72.

Snee, R. D., R. W. Hoerl and A. N. Patterson (2008) “In with The Right Crowd; Getting Management on Board to Support Statisticians’ Roles and Initiatives”, Quality Progress, May 2008, 70-73.



Questions and Comments?

For Further Information, Please Contact:

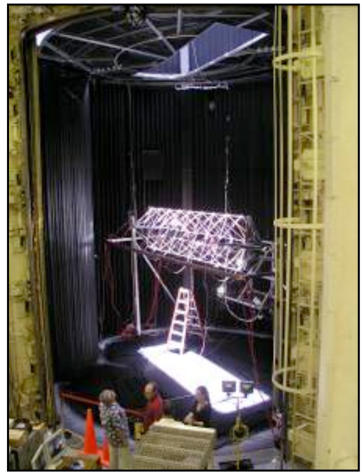
Ronald D. Snee, PhD

Snee Associates, LLC

Ron@SneeAssociates.com

610-213-5595

www.SneeAssociates.Com

NASA Engineering and Safety Center (NESC) Overview

May 5, 2011




NESC Background and Vision

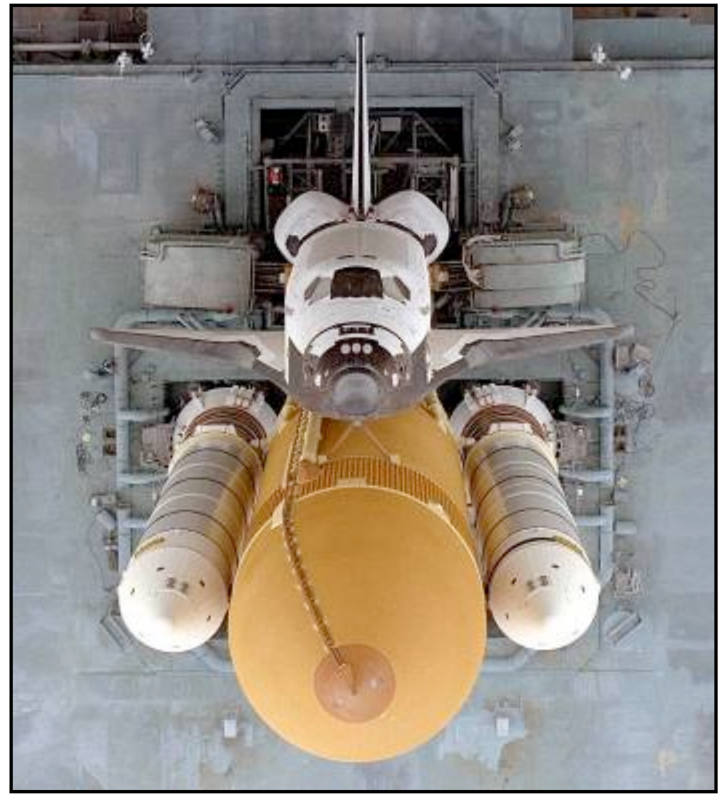
Apollo Saturn V Launch Vehicle

- NESC was established in July 2003 in response to the Columbia accident
- Safety philosophy has 3 tenets:
  - Strong in-line checks and balances
  - Healthy tension
  - "Value added" independent assessment
- NESC provides independent assessment of technical issues for NASA programs and projects

NESC is cultivating a Safety culture focused on **engineering and technical excellence**, while fostering an **open environment** and attacking challenges with **unequalled tenacity**

NESC Model

- Institutionalized "Tiger Team" approach to solving problems
- Agency-recognized NASA Technical Fellows lead Technical Discipline Teams (TDT)
  - TDTs include "ready" experts from across NASA, industry, academia and other government agencies
- Assemble diverse, expert technical teams that provide robust technical solutions to the Agency's highest-risk and most complex issues
- Strong Systems Engineering function for proactive trending and identification of problem areas before failures occur



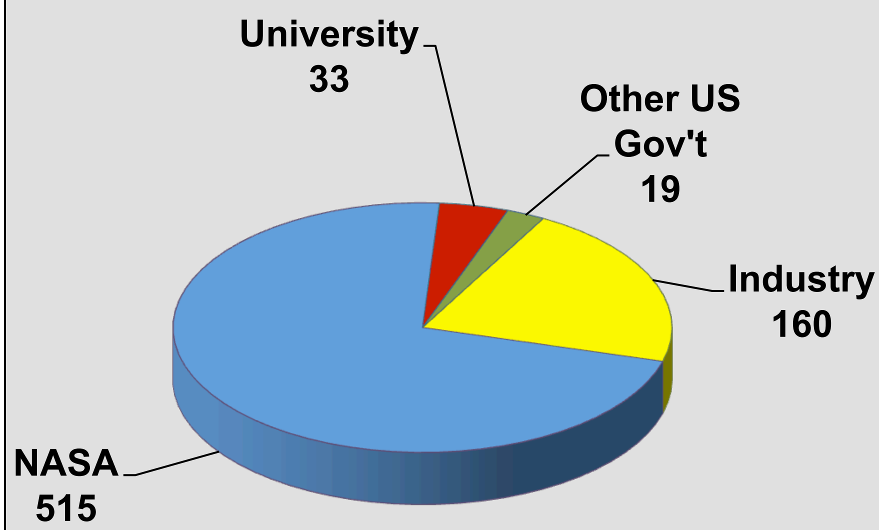
Space Shuttle on Mobile Launch Platform

Focus on technical rigor and engineering excellence

NESC Organization

Distributed NESC Team

- NESC has 61 full-time employees selected from across the Agency and externally
- NESC Chief Engineers at each Center provide technical insight and liaison roles
- 15 NASA Technical Fellows are recognized experts in their respective engineering fields



Over 700 top caliber engineers support the NESC nationwide

- 18 Technical Discipline Teams (TDT) comprised of 16 engineering and 2 operations disciplines create a network of over 700 engineers available for matrix support
- More than 200 TDT members are drawn from industry, academia and other government agencies

NASA Technical Fellows Disciplines

- 15 NASA Technical Fellows are currently active
  - Aerosciences - Dave Schuster (LaRC)
  - Avionics - Oscar Gonzalez (GSFC)
  - Electrical Power – Denney Keys (GSFC)
  - Flight Mechanics – Dan Murri (LaRC)
  - Guidance, Navigation, and Control - Neil Dennehy (GSFC)
  - Human Factors - Cynthia Null (ARC)
  - Life Support / Active Thermal - Hank Rotter (JSC)
  - Loads and Dynamics - Curt Larsen (JSC)
  - Materials - Bob Piascik (LaRC)
  - Mechanical Systems – Joe Pellicciotti (GSFC)
  - Non-Destructive Evaluation - Bill Prosser (LaRC)
  - Passive Thermal – Steve Rickman (JSC)
  - Propulsion – Roberto Garcia (MSFC)
  - Software - Mike Aguilar (GSFC)
  - Structures - Ivatury Raju (LaRC)

- Four additional disciplines to be added pending available funding
  - Space Environments
  - Systems Engineering
  - Cryogenics
  - Instruments and Sensors

NESC Resident Engineer Program

- Creates an opportunity to allow early career participants to gain hands on experience
  - Provide a technically diverse learning experience for resident engineers within the NESC organizational framework
  - Gain first-hand experience working with NASA technical experts and leaders
- Builds upon the principles of the MLAS Resident Engineer model
- Benefits the Agency by connecting to a younger generation and providing a fresh perspective to technical activities
- One year detail assignment for GS-12's and 13's

2009–2010 NESC Resident Engineers

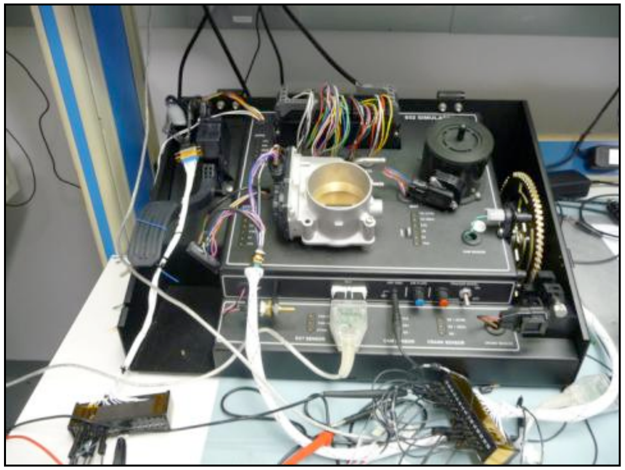
2010-2011 NESC Resident Engineers

NESC Technical Highlights

Shell Buckling Knockdown Factor Testing

Crew Module Water Landing Modeling Assessment

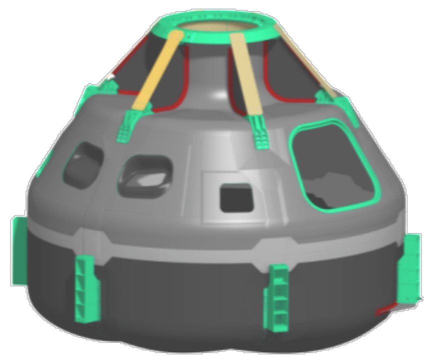
NESC Technical Highlights



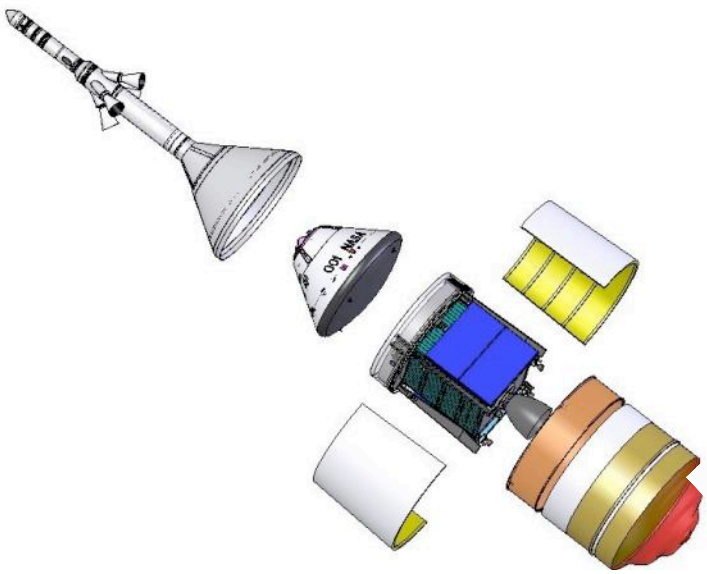
NHTSA Toyota Sudden Acceleration Investigation

Support for Chilean Miners Rescue Effort

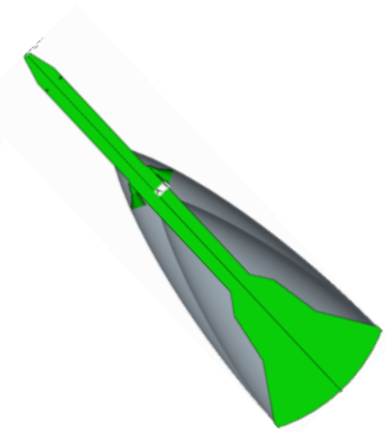
NESC Leading Agency-Wide Teams Gaining Hands-On Experience In Design, Development, and Test



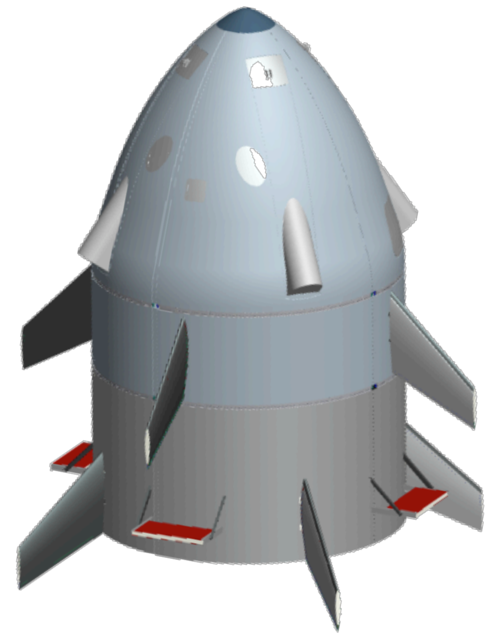
Composite Crew Module



CEV Smart Buyer



Alternate Launch Abort System



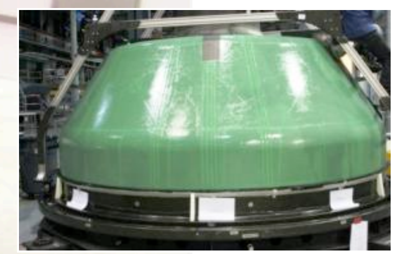
Max Launch Abort System

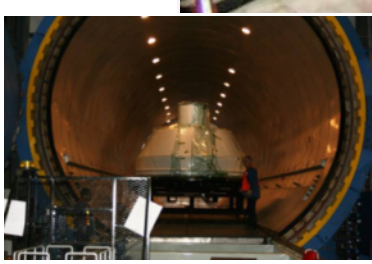
NESC Leading Agency-Wide Teams
Max Launch Abort System

Develop an alternate launch abort system design as risk mitigation for the Orion LAS and demonstrate the concept with a pad abort flight test.

NESC Leading Agency-Wide Teams
Composite Crew Module

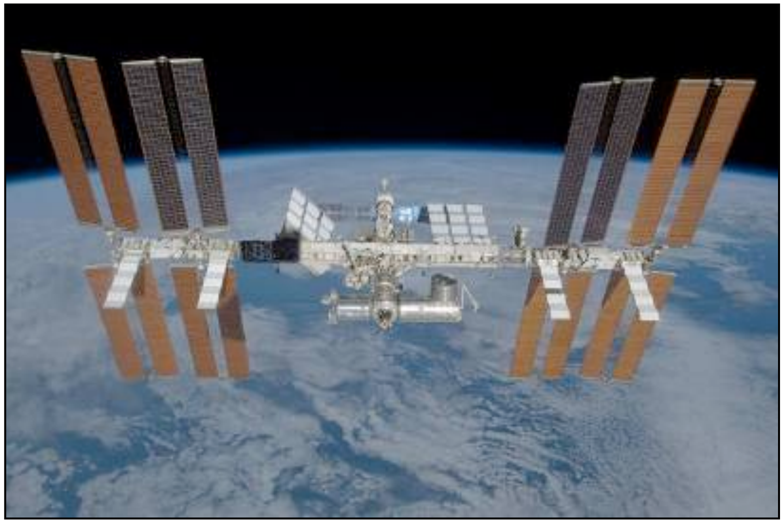
Design and build a composite crew module to gain hands-on design, build, and test experience.





Contributions to the Agency

- After 7 years and 400+ technical assessments the NESC has become the "value added" independent technical organization for the Agency
- The NESC model provides an excellent example of the benefits of bringing together diverse technical experts to solve the Agency's most difficult problems
    - Creative, robust technical solutions
    - Stronger checks and balances
    - Well informed decision making
- The NESC has fulfilled a role for off-line design, development and test to provide alternate solutions, gain valuable hands-on experience, and help train the next generation of engineers



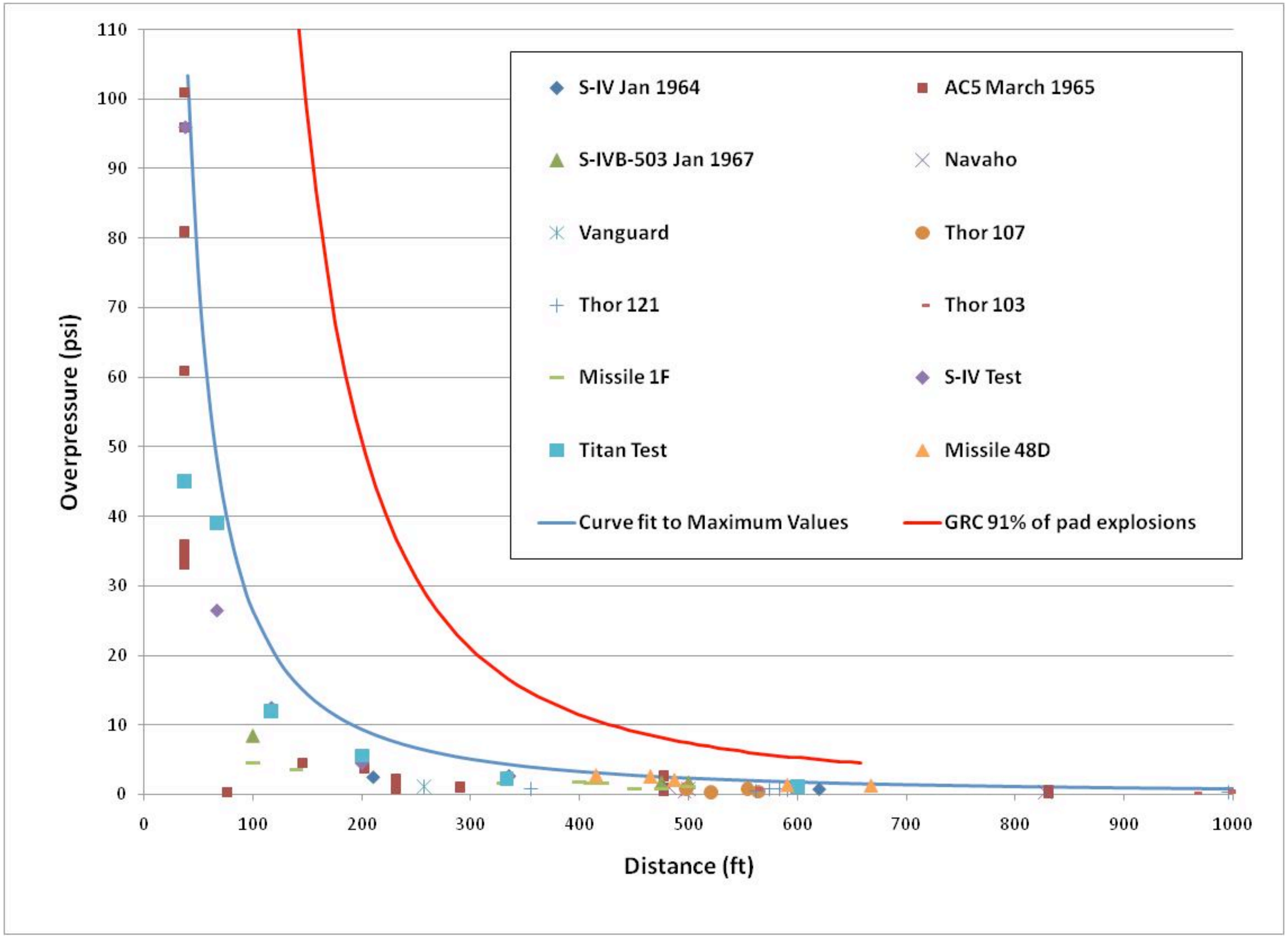
A Challenge for Statistical Engineering Community: Leveraging Limited Data

Case Study: Launch Environments

- Problem
  - Catastrophic failures of launch vehicles during launch and ascent are currently modeled using equivalent TNT estimates
  - This approach tends to over-predict the blast effect with subsequent impact to launch vehicle and crew escape requirements
  - Work has begun on a less-conservative model based on historical failure and test data coupled with physical models and estimates

- Challenge
  - Revised approach requires a statistical assessment of historical databases
  - NESC was asked to conduct a peer review of the work and provide findings and recommendations

Case Study: Launch Environments



The Challenge

- Challenge: Agency senior leaders need timely inputs to make well informed decisions, often with limited data.

- What can the statistical engineering community do to help?
  - What tools/methods are available or can be developed to deal with small data sets for real-time problem solving?
  - How can statistical information be communicated more clearly to decision makers outside of the community?
  - How can the community engage earlier in issue resolution to help guide the testing and data collections?

Truths, Darn Truths, and Statistics

Dr. Ed Kraft AEDC/CZ

Presented at the

NASA Statistical Engineering Symposium

Williamsburg, VA

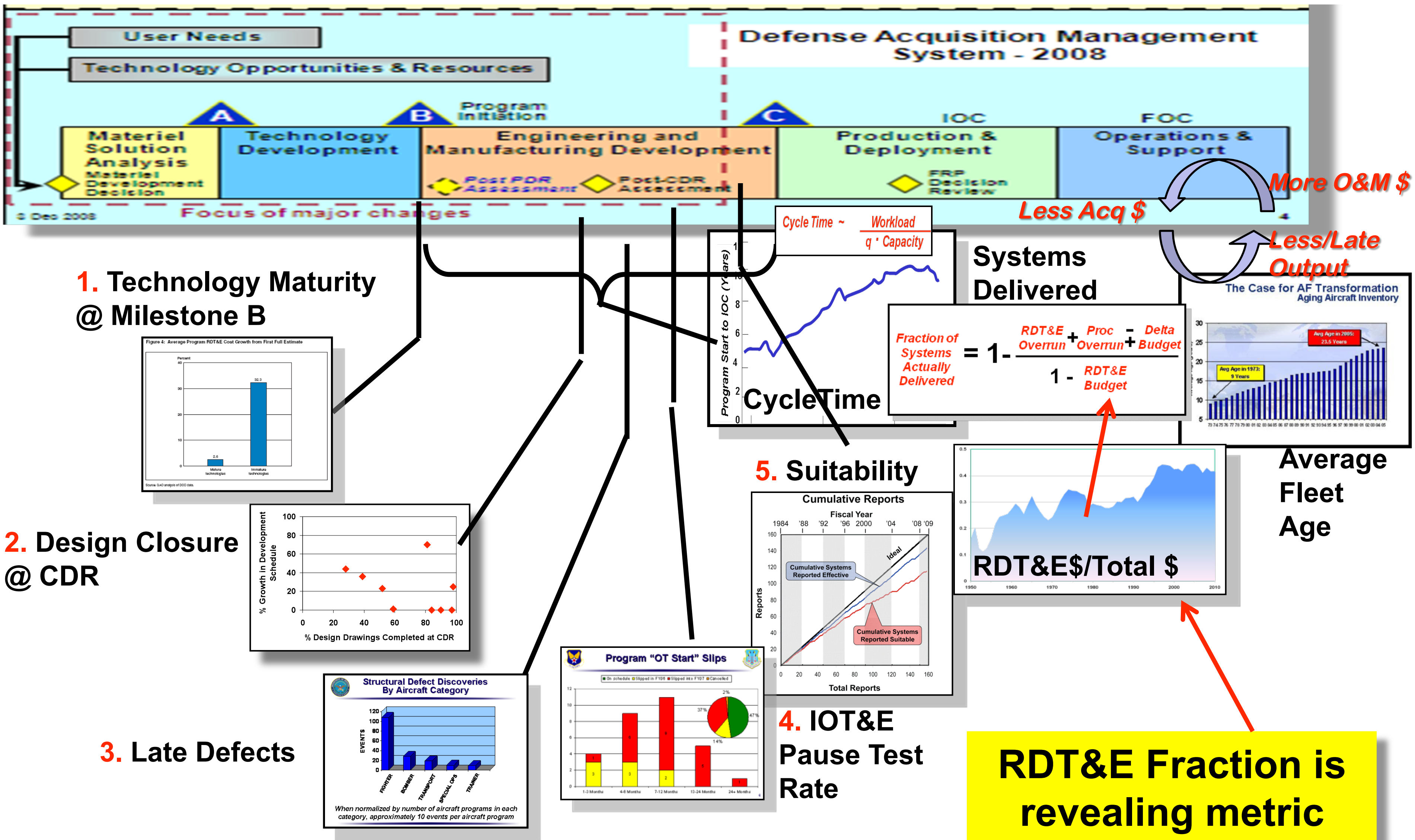
3-5 May, 2011

Objective

- Shift image from lies, damn lies, and statistics to
  - Truth – right answer achieved in an experiment
  - Darn Truth – defined risk at key leverage points in acquisition
  - Statistics – disciplined approach to identifying and managing risk

It is less about statistical theory and more about how we increase systems engineering effectiveness during defense acquisition using statistical tools

Targeting Five Key Leverage Points
Marked by Events – Mired by Lack of Effectiveness



Acquisition Cycle Time
Key T&E Effectiveness Parameter

$$\text{Cycle Time} \sim \frac{\text{Workload}}{q \cdot \text{Capacity}}$$

- *Workload* – Process driven, currently ~22,000 of wind tunnel testing and 13,000 of propulsion cell testing
- *q (inverse of rework)* – Process driven, typically have 10 structural failures found in flight; control surface resizing
- *Capacity* – Budget driven, availability x staffing x throughput

50% reduction in wind tunnel costs equates to just a few tenths of a percent reduction in program costs – Reducing acquisition cycle time by a month could save more than the cost of the entire wind tunnel campaign

Kraft, Edward M. " Integrating Computational Science and Engineering with Testing to Re-engineer the Aeronautical Development Process," AIAA Paper 2010-0139 , January, 2010.

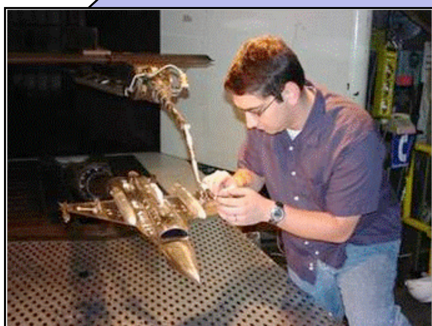
Kraft, Edward M. and Huber, Arthur F II "A Vision for the Future of Aeronautical Ground Testing," The ITEA Journal of Test and Evaluation, Vol 30, No 2, June 2009.

Streamlining Testing at the Campaign Level
New T&E Tools + DOE

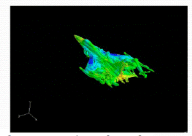
AFMC

Common Thread System ID Techniques

"Fly the Mission" Ground Testing



Computational Science and Engineering Dynamic Trajectories



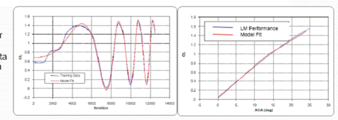
Computational Research and Engineering for Acquisition Tools and Environments - Air Vehicles

Kestrel

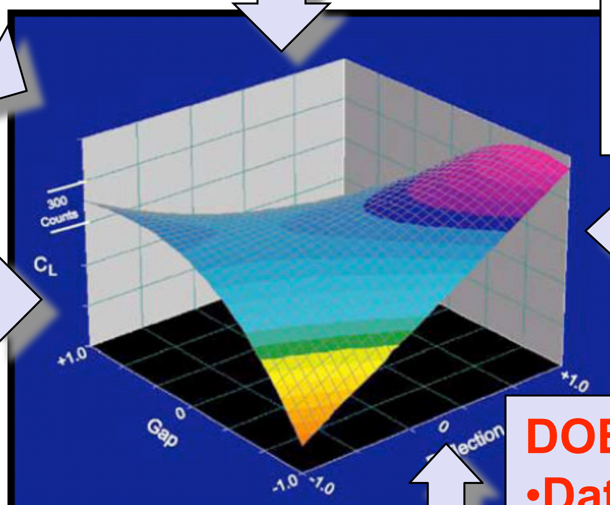
Kestrel v1.0 Results

Capability 2
Rigid Body Prescribed Motion

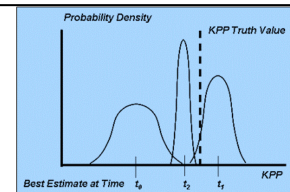
- Pitch Training Maneuver
  - M=0.6, Alt.=10k
  - Varying Alpha amplitude and frequency
  - Compute $C_L(t)$
- Build a SIDPAC Model: $C_L(\alpha,q,\dot q) = C_0 + C_1\alpha + C_2 q + C_3 q^2\alpha + C_4\dot q\alpha + C_5 q^4 + C_6\dot q q^2 + C_7 q\alpha^2 + C_8\dot q q + C_9\alpha^3 + C_{10}\dot q + C_{11}\dot q^3 + C_{12}\dot q^2 + C_{13} q^2 + C_{14} q\alpha$
- SIDPAC model gives…
  - dynamic behavior for **ANY** maneuver inside the regressor space
  - static coefficient data CL, CD, CM –alpha curves **FREE**



Flight Testing



Estimation Theory Quantify Effectiveness of Testing



Using Estimation Theory* variance reduction is proportional to the effectiveness of resources used and resources applied

$p(t_{i+1}) = p(t_i)/(1 + p(t_i)\ u\ \Delta t)$, u = resource effectiveness
Or
$u(t) = (p(t_i)/p(t_{i+1})-1)/p(t_i)\ \Delta t$

Which can be estimated used the SEMP, TEMP, and KPP values pre- and post-test

Value of T&E

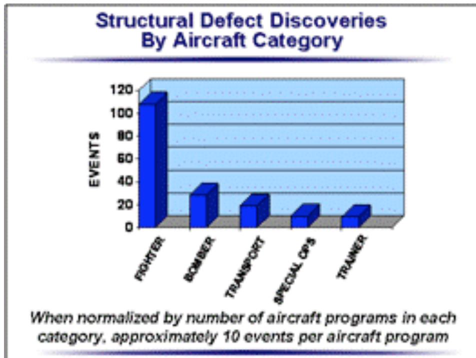
DOE
- **Data Merge/Data Mine**
- **Response Surface Analysis**
- **Variance Reduction Strategy**

Kraft, Edward M. " Integrating Computational Science and Engineering with Testing to Re-engineer the Aeronautical Development Process," AIAA Paper 2010-0139 , January, 2010

Reducing Late Defect Discoveries
Bayesian Statistics + DOE

Increasing Costs of Defects

Bayesian Parameter Estimation for Risk Assessment



Effect 2
Structural Flaw
Found In Flight

$P(E_2|E_1)$

DOE Power Coefficients?

$P(C_4)$

Cause 4
Manufacturing/
Assembly Flaw

$P(E_1|C_4)$

Effect 1
Prototype
System Flaw

$P(E_1) = \Sigma\ P(C_n) \cdot P(E_1|C_n)$

$P(E_1|C_1)$

$P(E_1|C_3)$

$P(E_1|C_2)$

Cause 1
Design Flaw

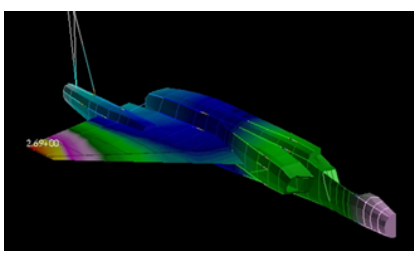
$P(C_1)$

Cause 2
Analysis
Flaw

$P(C_2)$

Cause 3
Ground Test
Flaw

$P(C_3)$



Bayesian Learning Model?

New Information
Model/Data

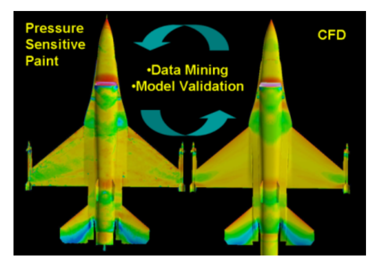
Current
Knowledge

$P(y|\theta)$

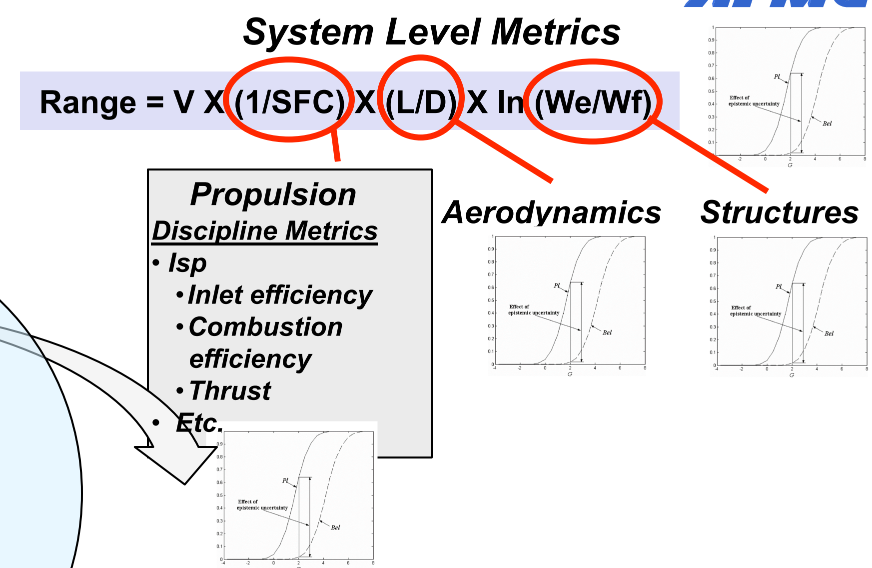
Updated
Knowledge

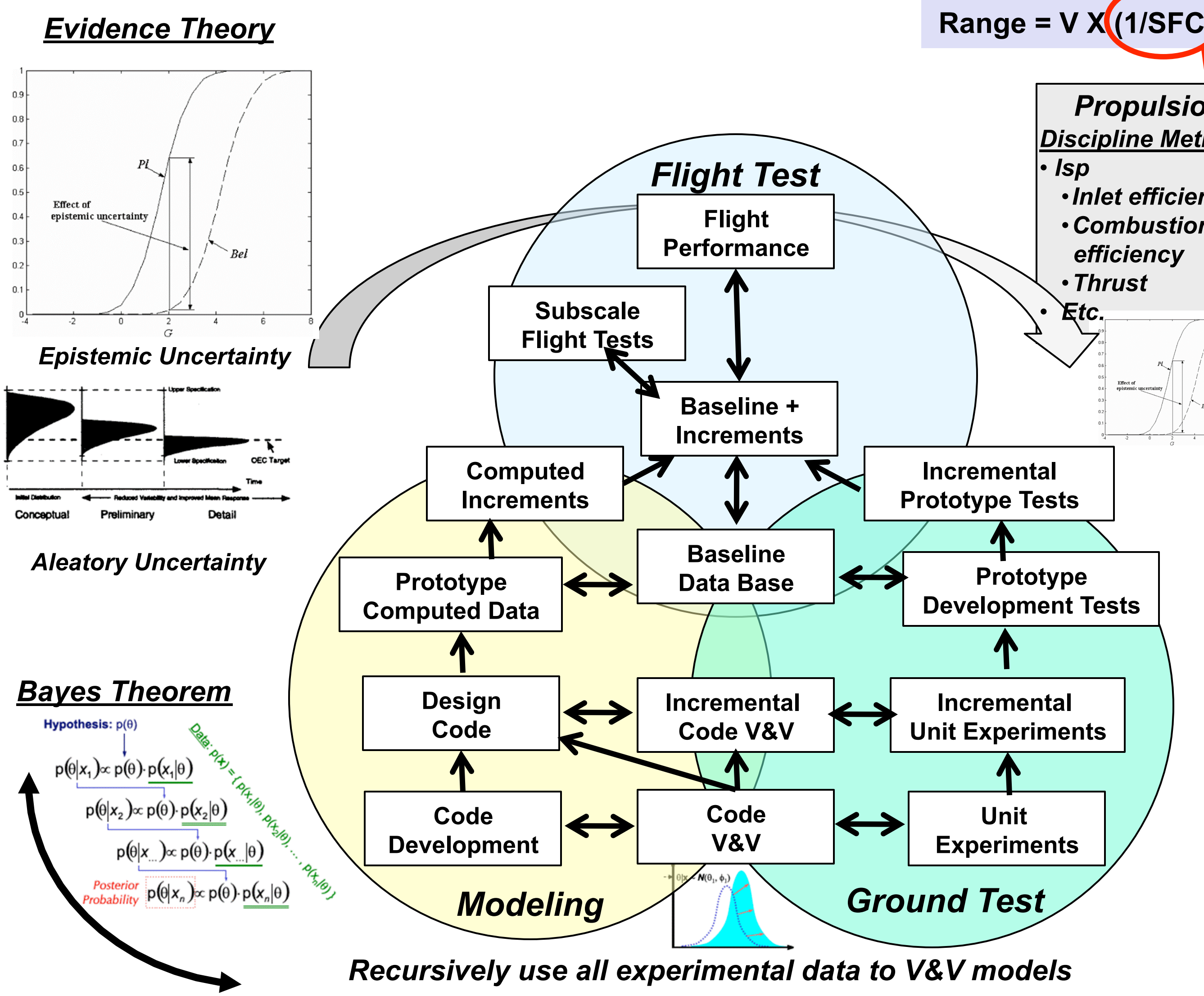
$P(\theta)$

$P(\theta|y)$



Probabilistic Incremental Development
Merging Modeling and Testing





Adapted/Merged from:

- Kraft, E. M., Chapman, G., "A Critical Review of the Integration of Computations, Ground Tests, and Flight Test for the Development of Hypersonic Vehicles," AIAA-93-51 01, Munich, Germany, Nov. 30-Dec. 3, 1993
- Mantis, G.C., Mavris, D. "A Bayesian Approach to Non-Deterministic Hypersonic Vehicle Design," AIAA Paper 2001-01-3033, 2001.
- Roy, C.J., Oberkampf, W.L. "A Complete Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing," AIAA Paper 2010-124, Presented at the 48th AIAA ASM Conference, Orlando, FL, January 2010.

Summary

- Statistics and DOE is an increasingly important tool in T&E – best current value is *a priori* definition of data quality required
- T&E Community gaining expertise in DOE
- Applications of DOE in isolated tests currently produces marginal return to a program
  - Not pre-planned in TEMP
  - Inertia of conservative, legacy processes
- Integration of DOE with new T&E tools, Estimation Theory, and Bayesian Logic could amplify impact of DOE
- Best ROI for DOE needs to be based on acquisition program cycle time

Statistical Research – Fostering Mutually Beneficial Collaborations

Christine M. Anderson-Cook, Ph.D.
Statistical Sciences Group (2004-present),
Los Alamos National Laboratory

Previously, Faculty member at Virginia Tech (1996-2004)

History of LANL Collaboration with Academia

- Visiting Faculty Program in Statistical Sciences Group at LANL for many years with visitors from many universities
- We recognized the need for expertise from university faculty to:
  - Broaden our expertise base
  - Broader perspective of where our problem fits into existing work
  - Provide validation and feedback for new research
  - Opportunity to publish
- Characteristics of LANL statistics work:
  - Need "good solutions" in a timely fashion (sometimes "ideal solution" not available quickly enough)
  - Reoccurring themes and problems
  - Nearly endless supply

LANL Statistics Group and ISU Department of Statistics received the ASA SPAIG (Statistical Partnerships between Academia, Industry and Government) Award for our multi-year collaboration

What is a Successful Academic-Industry-Government collaboration?

- From the Academic point of view
  - Funding (summer salary and metric of success for universities)
  - Good problems and data
  - Publications
  - Experience for graduate students and faculty
- From the Industry and/or Government point of view
  - Access to expertise
  - Availability of time to work on problems
  - Excellent solutions
  - Publications (sometimes yes, sometimes no)

Areas of Collaboration:
Design and Analysis of Experiments

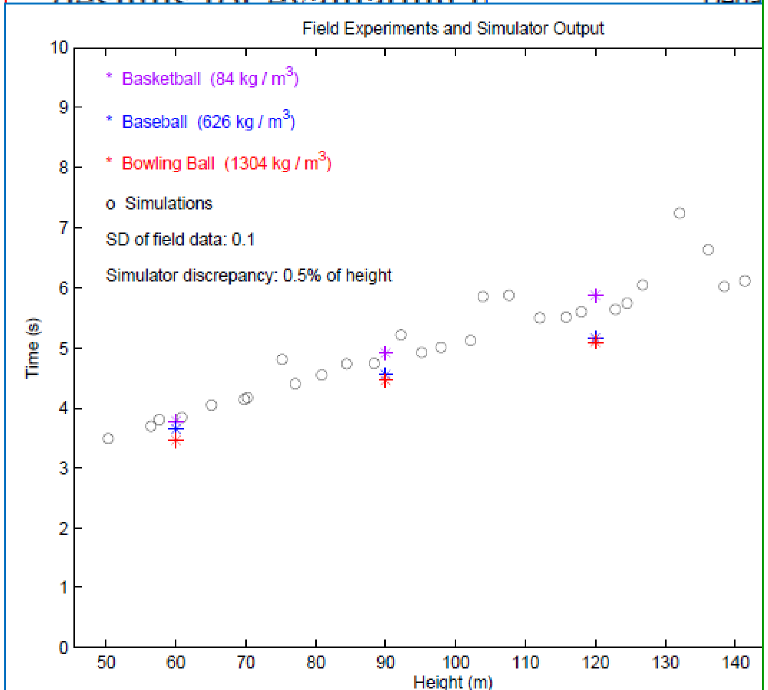
Available online at www.sciencedirect.com
SCIENCE DIRECT
journal of
Journal of Statisti
Inference 136 (20

ELSEVIER

Sampling plans based or
designs for evaluating t

Using Orthogonal Arrays in the Sensitivity Analysis of Computer Models

Max D. MORRIS
Leslie M. MOORE and Michael D. MCKAY
Depa



Integrated Analysis of Computer and Physical Experiments

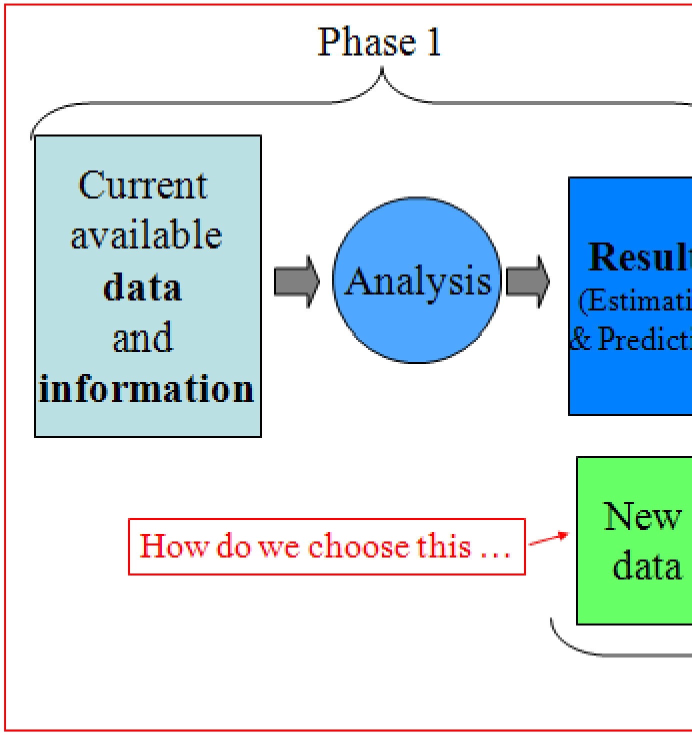
C. Shane REESE
Department of Statistics
Brigham Young University
Provo, UT 84602
(reese@stat.byu.edu)

Alyson G. WILSON, Michael HAMADA, and Harry F. MARTZ
Statistical Sciences Group
Los Alamos National Laboratory
Los Alamos, NM 87545

Kenneth J. RYAN
Department of Statistics
University of Illinois
Chicago, IL 60208

Scientific investigations frequently involve data from computer experiment(s) as well as related physical experimental data on the same factors and related response variable(s). There may also be one or more expert opinions regarding the response of interest. Traditional statistical approaches consider each of these datasets separately with corresponding separate analyses and fitted statistical models. A compelling argument can be made that better, more precise statistical models can be obtained if the combined data are analyzed simultaneously using a hierarchical Bayesian integrated modeling approach. However, such an integrated approach must recognize important differences, such as possible biases, in these experiments and expert opinions. We illustrate our proposed integrated methodology by using it to model the thermodynamic operation point of a top-spray fluidized bed microencapsulation processing unit. Such units are used in the food industry to tune the effect of functional ingredients and additives. An important thermodynamic response variable of interest, **Y**, is the steady-state outlet air temperature. In addition to a set of physical experimental observations involving six factors used to predict **Y**, similar results from three different computer models are also available. The integrated data from the physical experiment and the three computer models are used to fit an appropriate response surface (regression) model for predicting **Y**.

Areas of Collaboration: Resource Allocation



Computationally Efficient Evaluation of Resource Allocation Alternatives for Estimating the Reliability of Complex Systems

Jessica Chapman
St. Lawrence University
Canton, NY 13617

Max Morris
Iowa State University
Ames, IA 50011

Christine Anderson-Cook
Los Alamos National Laboratory
Los Alamos, NM 87545

July 6, 2010

Abstract

Resource allocation considers the problem of quantitatively determining, given a fixed amount of resources, how to best use those resources to collect additional data to supplement an initial analysis. We address resource allocation in the context of the second-stage in system reliability studies; this second-stage data is used to obtain a more precise estimate of the system's reliability. The current strategy for addressing the research allocation problem is computationally intensive and time-consuming. We present new, computationally efficient methodology that can be used to quickly assess the anticipated improvements for candidate allocations and demonstrate its effectiveness with a missile system application.

KEYWORDS: Resource Allocation, System Reliability, Bayesian Reliability, Markov Chain Monte Carlo, Experimental Design

Areas of Collaboration: Measurement Systems

Elementary Statistical Methods and Measurement Error

Stephen B. VARDEMAN, Joanne R. WENDELBERGER, Tom BURR, Michael S. HAMADA, Leslie M. MOORE, J. Marcus JOBE, Max D. MORRIS, and Huaiqing WU

How the sources of physical variation interact with a data collection plan determines what can be learned from the resulting dataset, and in particular, how measurement error is reflected in the dataset. The implications of this fact are rarely given much attention in most statistics courses. Even the most elementary statistical methods have their practical effectiveness limited by measurement variation; and understanding how measurement variation interacts with data collection and the methods is helpful in quantifying the nature of measurement error. We illustrate how simple one- and two-sample statistical methods can be effectively used in introducing important concepts of metrology and the implications of those concepts when drawing conclusions from data.

KEY WORDS: Accuracy; Bias; Calibration; Data collection; Linearity; Precision; Repeatability; Reproducibility; Statistical education.

Graduate Lecture Notes on Statistics and Measurement in Engineering and Physical Science*

Steve Vardeman, Huaiqing Wu, and Max Morris
Iowa State University

Mike Hamada, Joanne Wendelberger, Tom Burr, and Lisa Moore
Los Alamos National Lab

Marcus Jobe
Miami University of Ohio

July 12, 2010

Abstract

This outline summarizes the main points of a series of graduate lectures on statistics and measurement prepared at Iowa State University.

Areas of Collaboration: Reliability

A Prediction-Based Model Selection Approach

Adam Pintar
Department of Statistics
Iowa State University
Ames, IA 50011
(apintar@iastate.edu)

Christine M. Anderson-Cook
Statistical Sciences Group
Los Alamos National Laboratory
Los Alamos, NM 87545
(c-and-cook@lanl.gov)

Huaiqing Wu
Department of Statistics
Iowa State University
Ames, IA 50011

models
st that
e been
selects
of in-
etailed
sed on
h cost
strated
mance.
oduces

A Bayesian Approach to
Industrial Experiment

Bria
Mic
[1]Dep
Iowa
Ame
[2]Sta
Los
Los

Research Article

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1109 Published online in Wiley InterScience

Using Age
of Reliabil
from a Fir

Lu Lu and Christ

For single-use non-repai
effective management o
to predict the reliability
reliability, the age of the
In this paper we present
based on the currently kn
based on observed usage
to answer the questions
or 'For a given system wi
with associated uncertai
selected from the popula
the estimated probabilit
demographics of the pop
in 2010 by John Wiley &

Keywords: population sur

Choosing a Consumption Strategy for A Population of Units based on Reliability

Lu Lu[1], Christine M. Anderson-Cook[1], and Alyson G. Wilson[2]
[1]Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
[2]Iowa State University, Ames, Iowa 50011, USA

Abstract: Managers and decision makers are often faced with difficult decisions balancing multiple competing objectives when selecting between several strategies for how to use the units in their inventory or stockpile. This paper considers how to define different metrics which appropriately summarize the objectives of a good strategy, how to consider what impact unanticipated changes in the future might have, and how to combine several criteria into a decision when no global winner is likely. This process is discussed in the context of maximizing the reliability of a population of single-use non-repairable units, such as missiles or batteries, which are being consumed (used and removed from the population) as they age over time.

Keywords: probit regression, decision making, multiple competing objectives, stockpile management

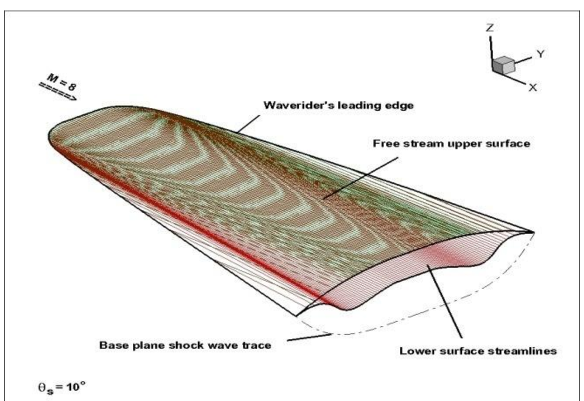
What makes it successful?

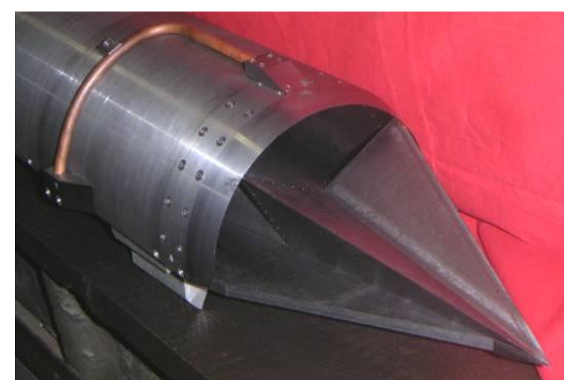
- The right people!!!
- Ongoing projects – spanning multiple years
- Good vision to identify reoccurring mission-critical areas (LANL) that are interesting, new and publishable (Academic partners)
- Natural match of expertise with problems
- Flexible timelines
- Opportunities for publications
- Funding not a major concern

Statistical Research - Industrial, Academic and Government Collaborations

Carolyn B. Morgan, Ph.D.
Professor, Department of Mathematics
Hampton University, Hampton, VA

<u>With significant engineering input</u>
<u>From Dr. Morris Morgan, III</u>





Previous Industrial and Academic Collaborations

- Statistician at GE Global Research (1972-1996)
  - ✓ Applied statistical methods to technically significant problems throughout the company (medical systems, GE Profile dishwasher design team, product reliability,..)
  - ✓ Collaborated with Dr. Morris Morgan, chemical engineering professor at Rensselaer Polytechnic Institute with

    13 years of prior experience with General Motors, Monsanto and GE Global Research and research projects with Aberdeen Proving Ground
  - ✓ Presented over 20 technical talks on statistical research work at ASA and AICHE national meetings and 15 technical papers

Current Industrial, Academic and Government Collaborations

- Transitioned to Hampton University in 1996
  - ✓Chair, Department of Mathematics
  - ✓Hampton Representative **to Universities Space Research Association** (collaborative membership organization of universities and other research organizations that cooperate with each other, with the United States government, and other entities to develop knowledge associated with space science and technology).
- Initiated interdisciplinary research with the School of Engineering and Technology to provide research opportunities, tuition and stipend support for STEM undergraduates and Applied Mathematics graduate students in the Statistics track

Statistical Research Presentations

Morgan, C. B. and Morgan, M. H. (2000). *A Statistical Model of the Drag Coefficient in an Engineering Transport System,* **Joint Statistical Meetings**, 2000 Indianapolis, IN (contributed paper).

Morgan, C. B. and Morgan, M. H. (2001). *Modeling the Effect of Cycle Time Distribution on System Performance,* **Joint Statistical Meetings**, 2001 Atlanta, GA (contributed paper).

Morgan, C. B. and Morgan, M. H. (2002). *Time Series Analysis of a Closed-Loop Electro – Spouted Bed,* **Joint Statistical Meetings**, 2002 New York City (contributed paper).

Morgan, C. B. and Morgan, M. H. (2003). *Predicting the Onset of Flow Instabilities Using Time Series Methods,* **Joint Statistical Meetings**, 2003 San Francisco, CA (contributed paper).

Morgan, C. B. and Morgan, M. H. (2004). *Statistical Modeling of Flow Instabilities in an Engineering System,* **Joint Statistical Meetings**, 2004 Toronto, Canada (contributed paper).

Morgan, C. B. and Morgan, M. H. (2005). *Statistical Investigation of Chaotic Data Streams Using a Haar Wavelet Transform*, **ASA Joint Statistical Meetings**, 2005 Minneapolis, MN (contributed paper).

Statistical Research and Engineering Grant Activities

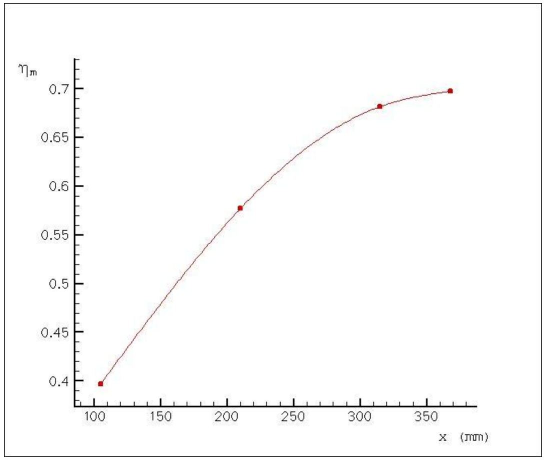
- NASA PACE-“MSET Educational Grant
  1997-1999, $300,000/3 yrs, Co-PI
- GE Fund- “GETMET” Educational Grant
  2000-2002, $150,000/2 yrs. Co-PI
- NASA- “Aero-Propulsion Center”
  2004-2008, $4,800,000/4yrs,
  Dr. Morris H. Morgan, III Center Director
- National Security Agency - “Statistical Data Mining and Analysis of Large Drifting Data Stream”
  2004-2006, $153,000/2yrs, Co-PI
- Lockheed-Martin & Orbital Sciences
  2010-2011, $400,00/1yr, Co-PI

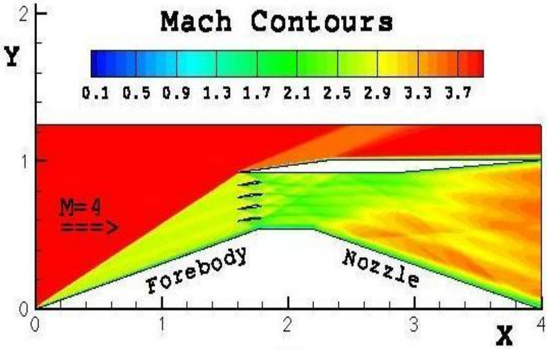
Hampton's Aero-Propulsion Center

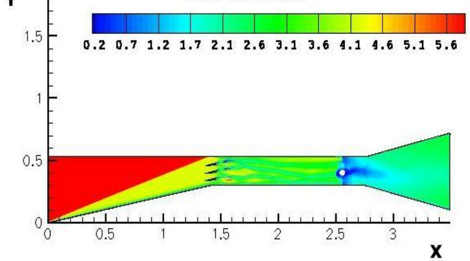
- Major NASA initiative to integrate statistical thinking, engineering and methods in aerospace engineering
- NASA funded project to advance research on
  - ✓ Improving Scramjet Combustion Efficiencies
  - ✓ Improving Lift to Drag Ratios of Wave-Rider Designs
  - ✓ Validating Numerical Simulations & Wind Tunnel Studies
  - ✓ Improving Signal Masking and Recovery
- Hampton Research Team

Engineering:	M. Morgan (Diretor), D. Lyons, J. Akyurtlu A. Akyurtlu & V. Khaikine
Statistics:	C. Morgan

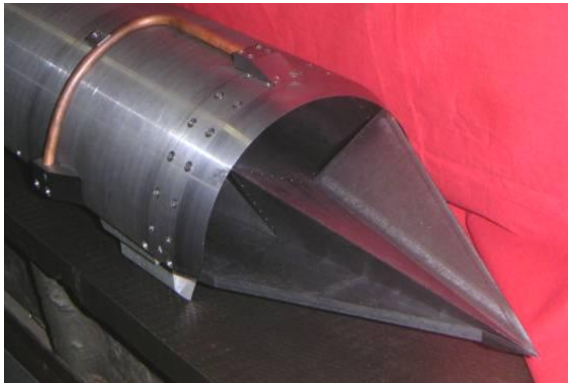
Improving Scramjet Combustion via Pylon Modifications

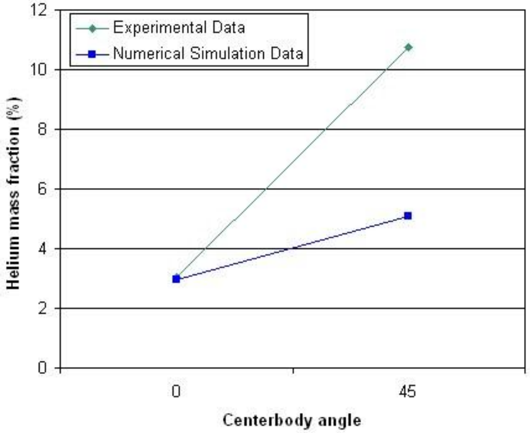


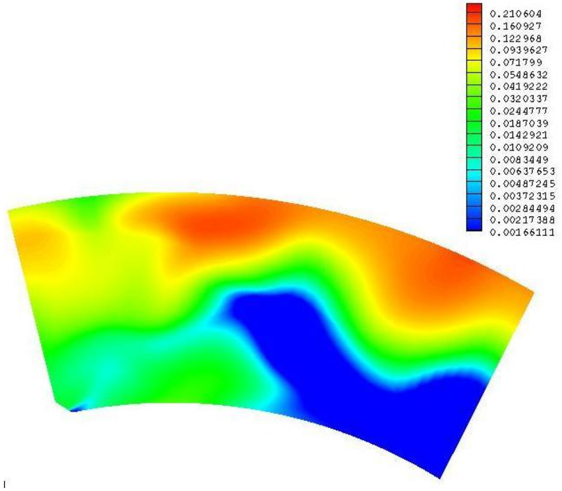




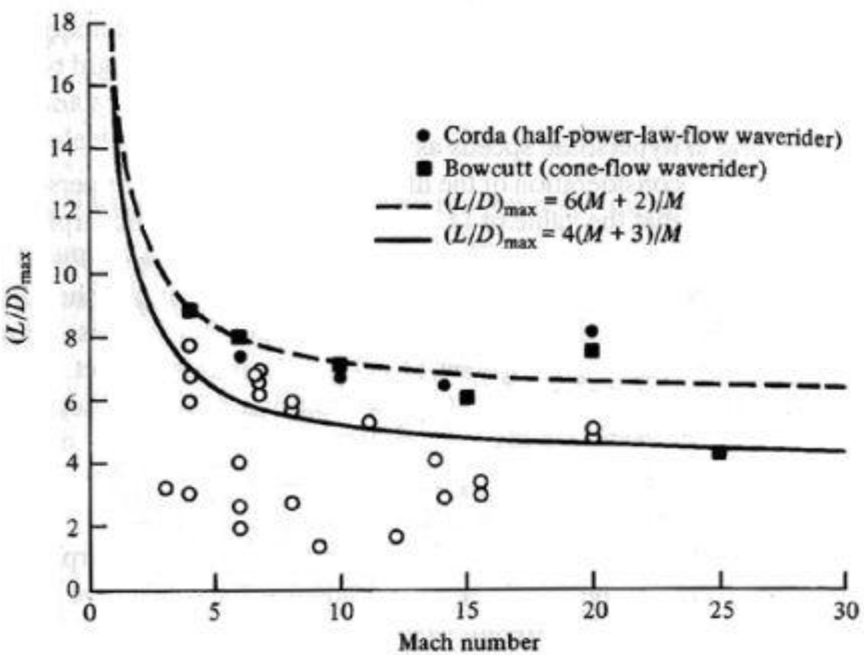
Star-Shaped Inlet Mixing Efficiencies

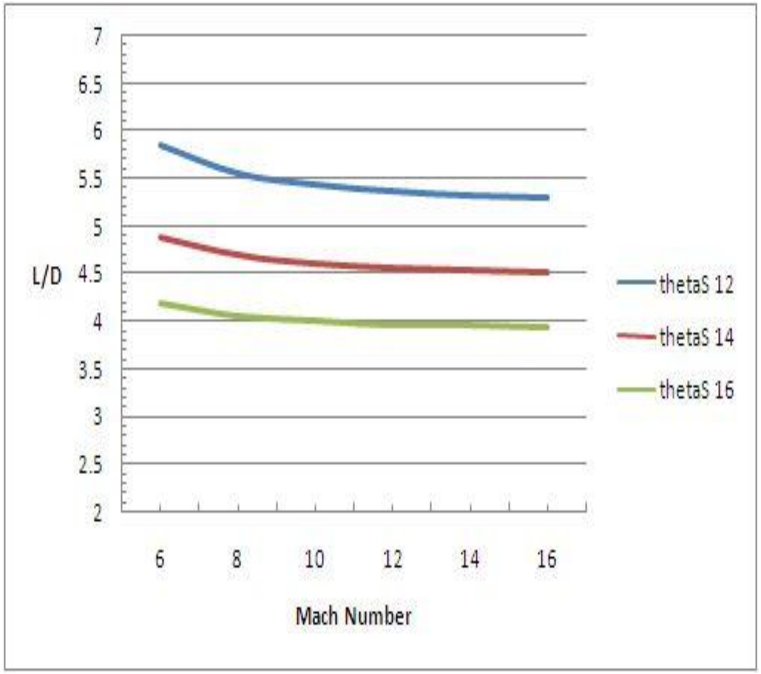






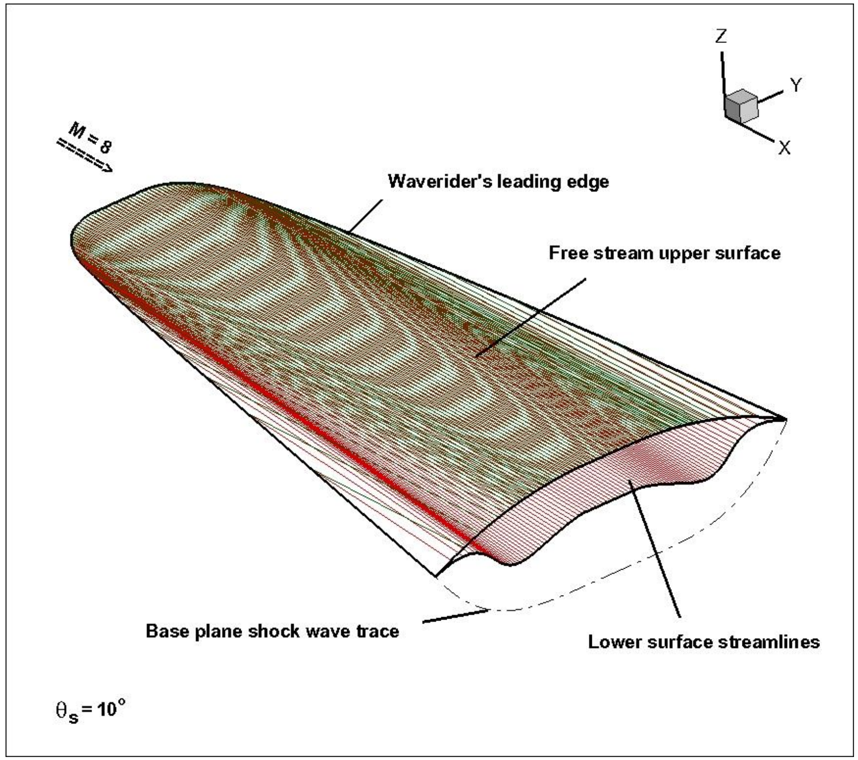
Added Dimensional Variable in Lift/Drag Ratio



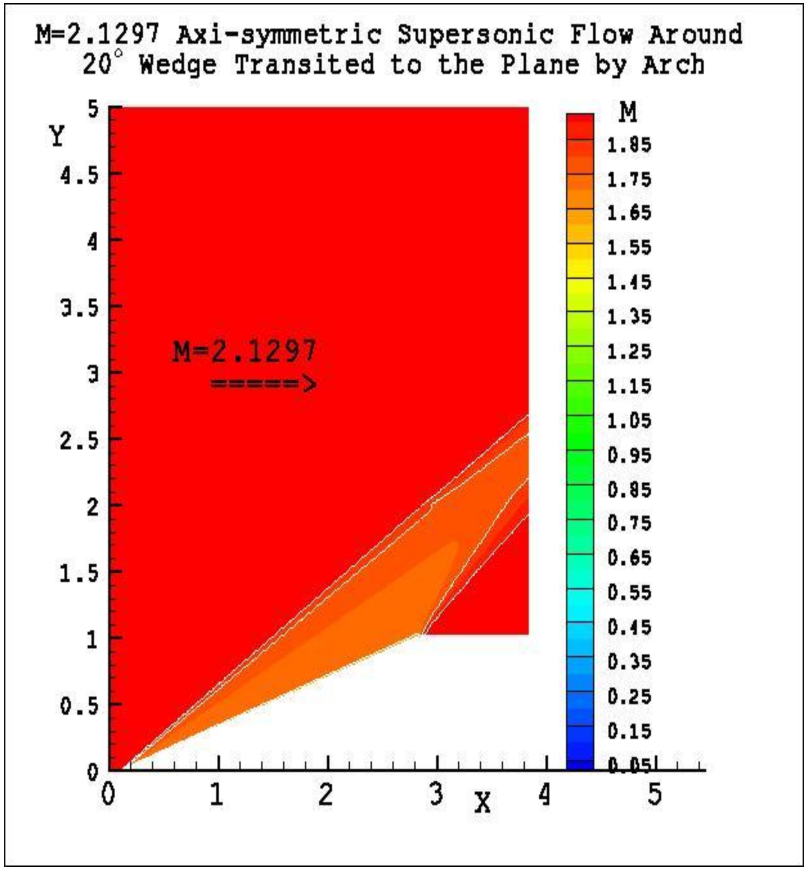


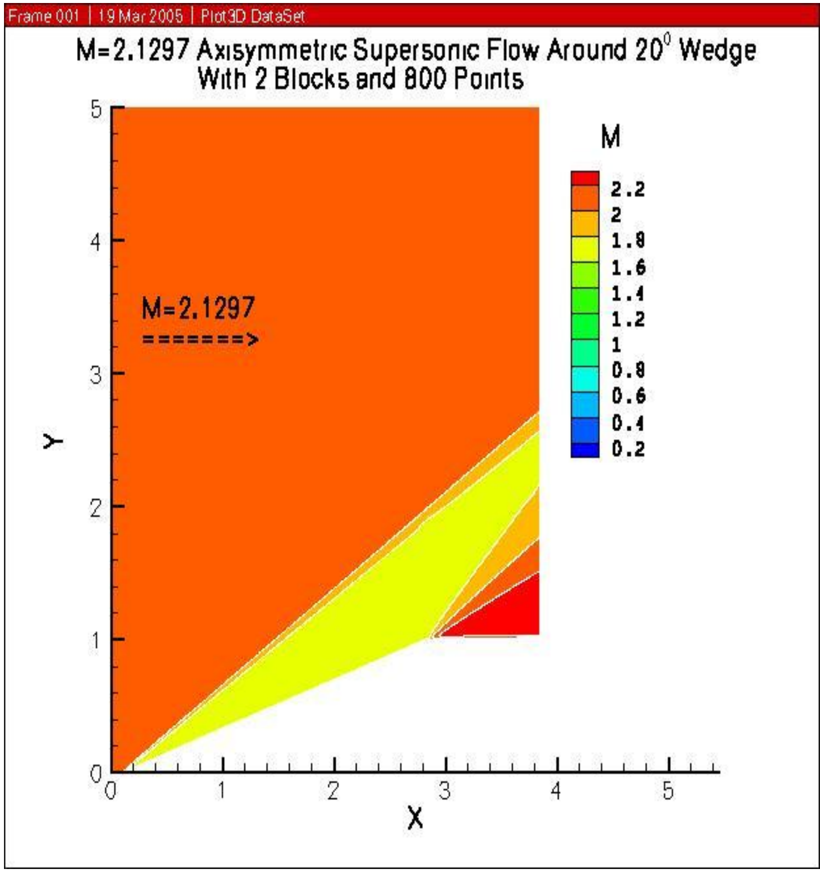
Lockheed Martin’s X-33 Design

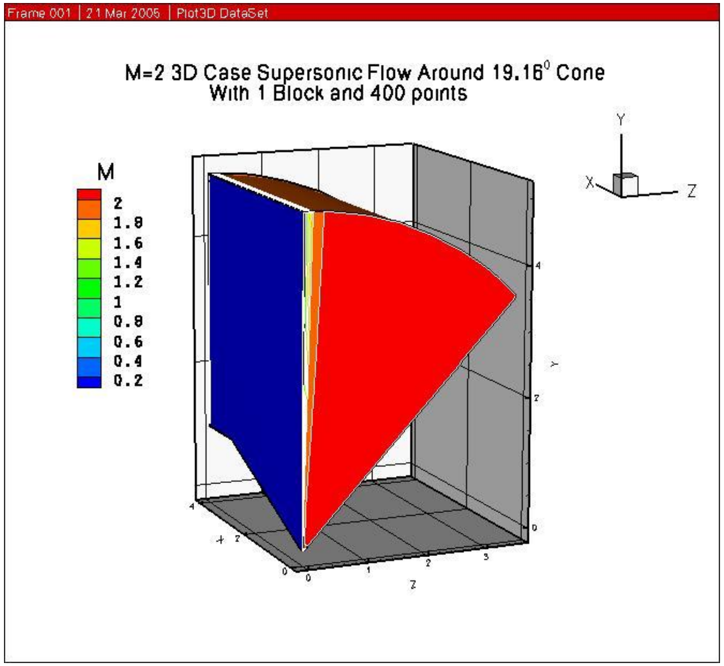
Improving Wave-Rider Designs

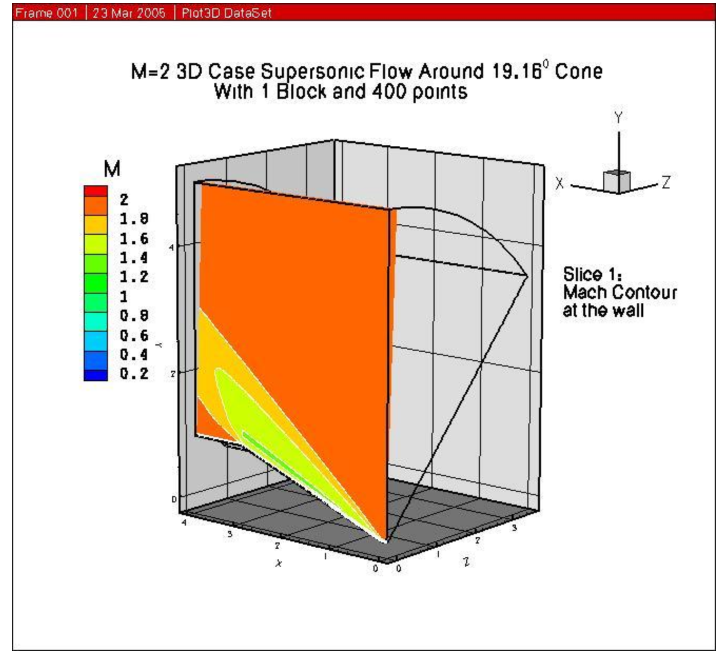


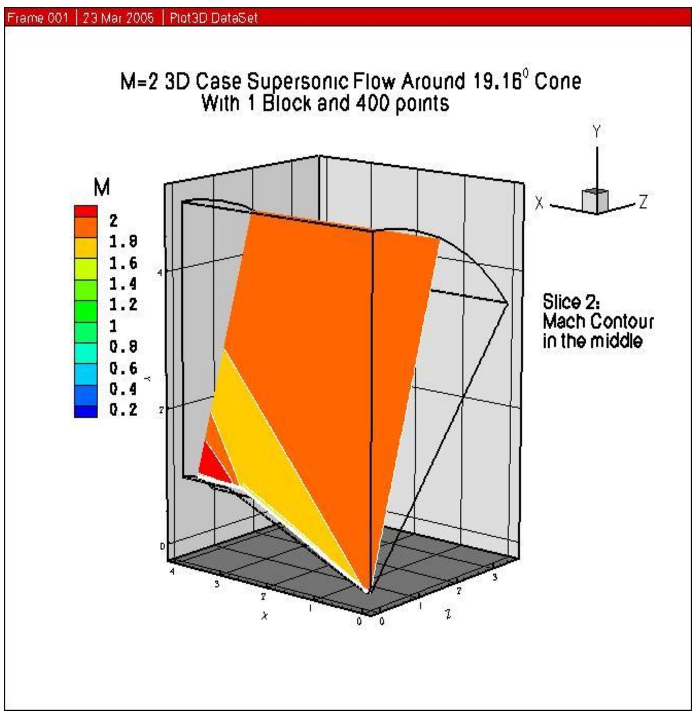
Validating Numerical Simulation & Wind Tunnel Studies

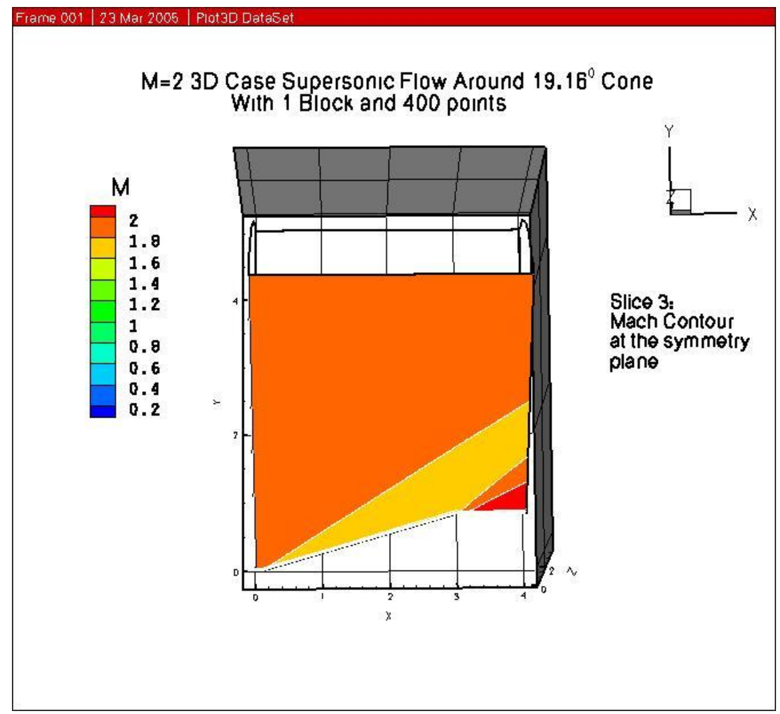





Frame 001 | 21 Mar 2005 | Plot3D DataSet
M=2 3D Case Supersonic Flow Around 19.16^0 Cone
With 1 Block and 400 points
M
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
Y
X
Z


Frame 001 | 23 Mar 2005 | Plot3D DataSet
M=2 3D Case Supersonic Flow Around 19.16^0 Cone
With 1 Block and 400 points
M
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
Y
X
Z
Slice 1:
Mach Contour
at the wall


Frame 001 | 23 Mar 2005 | Plot3D DataSet
M=2 3D Case Supersonic Flow Around 19.16^0 Cone
With 1 Block and 400 points
M
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
Y
X
Z
Slice 2:
Mach Contour
in the middle


Frame 001 | 23 Mar 2005 | Plot3D DataSet
M=2 3D Case Supersonic Flow Around 19.16^0 Cone
With 1 Block and 400 points
M
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
Y
X
Slice 3:
Mach Contour
at the symmetry
plane

Advantages of Industrial, Academic and Government Collaborations

- Key to addressing challenging problems that require interdisciplinary teams (engineers, statisticians, scientists, …) and large resources (equipment, tools, ..)
- Provides students and faculty “outside-the-book” research experiences
- Serves to attract future students to STEM careers
- Provides funding for undergraduate and graduate research and faculty
- Yields opportunities for publications, presentations, etc.

Statistical Research:
Some NASA Perspectives

Ray D. Rhew

NASA Langley Research Center

NASA Statistical Engineering Symposium

May 5, 2011

Opportunities and Challenges

- **Opportunities**
  - **As statistical thinking continues to be infused into NASA projects (Exploration, Science and Aeronautics)**
    - ➢ **New research opportunities will be revealed**
      - ➢ **Fundamental principles and theoretical derivations**
      - ➢ **Application of existing techniques to unique problems**
    - ➢ **Many avenues to obtain research capabilities**
      - ➢ **In-house personnel – stay current for future opportunities and collaborative efforts**
      - ➢ **Contracts – industry and academia**
      - ➢ **Research Announcements**
      - ➢ **Graduate research, post doctoral, undergraduate research**

Opportunities and Challenges

- Challenges
  - Educating organizations and projects on statistical engineering and its benefits
  - Clearly defining research opportunities
  - Securing resources within project environments (schedule and resource pressures)
  - Timely execution
  - Coordinating fiscal and academic calendars
  - Communicating benefits and applications

Statistical Engineering: What, Why and How

Panel Discussion

Moderator: Christine Anderson-Cook, LANL

Panelists:
Ronald Snee, Snee Associates
Geoff Vining, Virginia Tech
Mark Zabel, Straight Line Performance Solutions

NASA Statistical Engineering Symposium May 3-5, 2011

- What? (Led by Ron Snee)
  - Examples, Definition, as a Discipline
- Why? (Led by Mark Zabel)
  - Benefits to the individual statistician, team member and the organization
- How? (Led by Geoff Vining)
  - Roles and Challenges

Format: 30 min presentation, 15 min additional comments from other panelists, open discussion

- Founder and President of Snee Associates
- 24 years at Dupont
- Author of 3 books on Statistical Thinking and LSS with Roger Hoerl
- Academician in the International Academy for Quality, Fellow ASQ and ASA
- Recipient of ASQ Sherhart and Grant Medals
- Author or more than 225 papers in performance improvement, quality, management and statistics

- Professor and former Department Head in Department of Statistics at Virginia Tech
- Former editor of Journal of Quality Technology, and Quality Engineering
- 2010 ASQ Shewhart Medal recipient
- Winner of ASQ Brumbaugh and Lloyd Nelson award
- Author of 3 textbooks
- Internationally recognized expert in the use of designed experiments for quality and productivity improvement

- President and Co-founder of Straight Line Performance Solutions
- Statistical Engineer and Certified Master Black Belt
- Educated, mentored and coached more than 2250 LSS professionals
- 20+ years of experience in LSS, statistics, analytical methods, software development and process re-engineering
- Led design, redesign and improvement projects resulting in hundreds of millions of dollars of delivered value for a wide range of businesses

What is Statistical Engineering?

Ron Snee

Some Examples

- DMAIC Problem Solving and Improvement Framework
  - Links and sequences a variety of statistical and non-statistical tools
- DuPont Product Quality Management System
  - Integrates a variety of tools into a comprehensive quality system
- Strategy of Experimentation
  - System that links a variety of designs into an overall strategy
- GE Collections System
  - Process monitoring and improvement system that uses a variety of tools including flowcharts, control charts, Pareto charts and DOE
- Pharmaceutical and Biotech Quality by Design
  - Integration of process modeling, process and measurement control and process optimization creating process understanding

Major Impact is Created by Linking and Sequencing Many Different Tools into a System

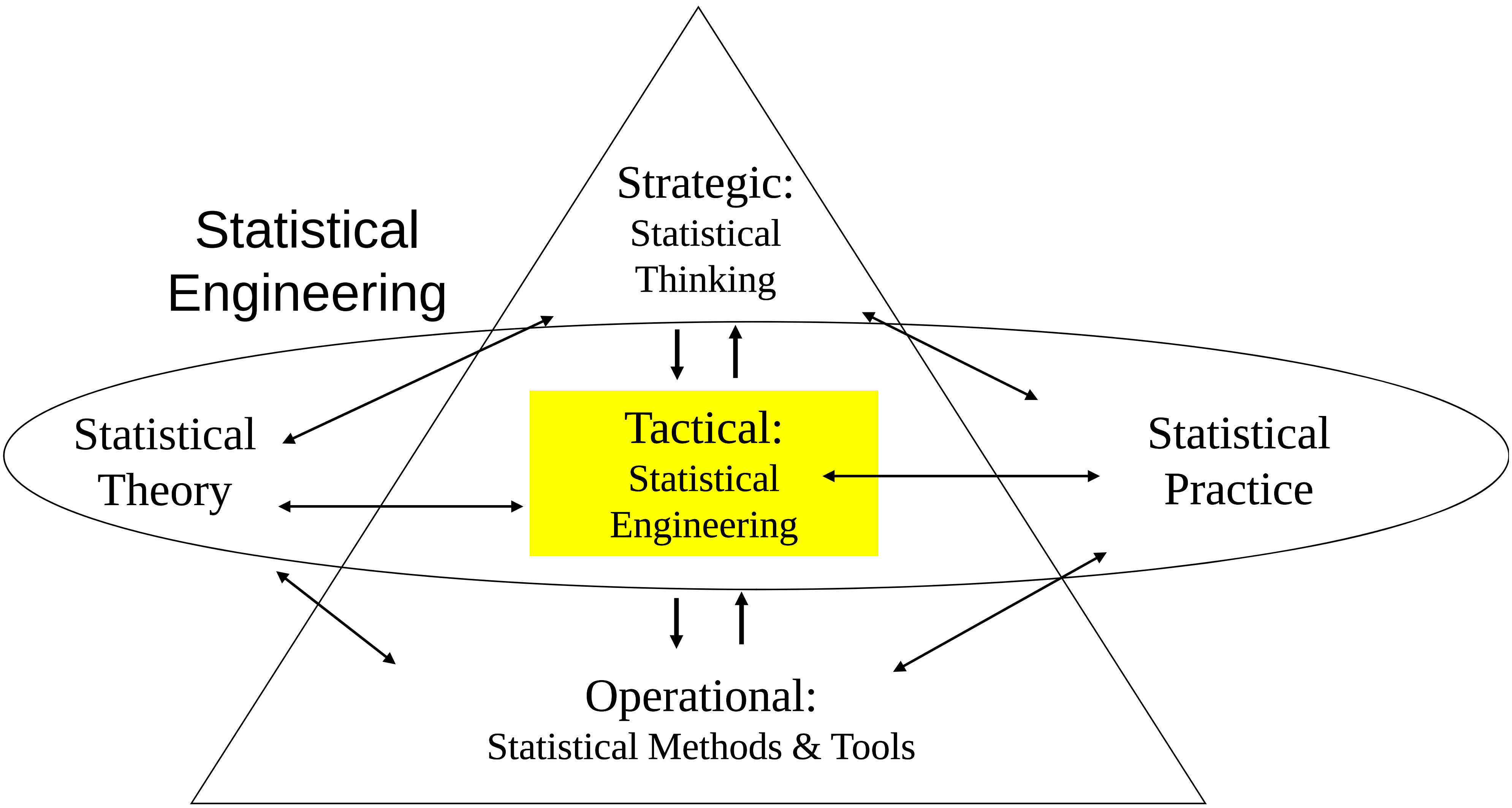
Statistical Engineering Definition

- Statistical engineering:
  - The study of how to best utilize statistical concepts, methods, and tools and integrate them with information technology and other relevant sciences to generate improved results (Hoerl and Snee 2010)
  - Trying to build something with the statistical science parts list

Notes

- For this to be a true engineering discipline as opposed to just a sexier term for applied statistics, there must be a dynamic theoretical foundation based on rigorous research, just as there is for electrical engineering, mechanical engineering, and so on
- This definition does not refer to application of statistics to engineering
  - Statistical engineering can be applied to improving anything
- This is a different definition than that used by Eisenhart, who we believe was the first to use this term in 1950

Statistical Engineering as a Discipline



Statistical Engineering - Discipline that Studies How to Drive Greater Impact from Existing Science and Theory of Statistics.

How Is It Different?

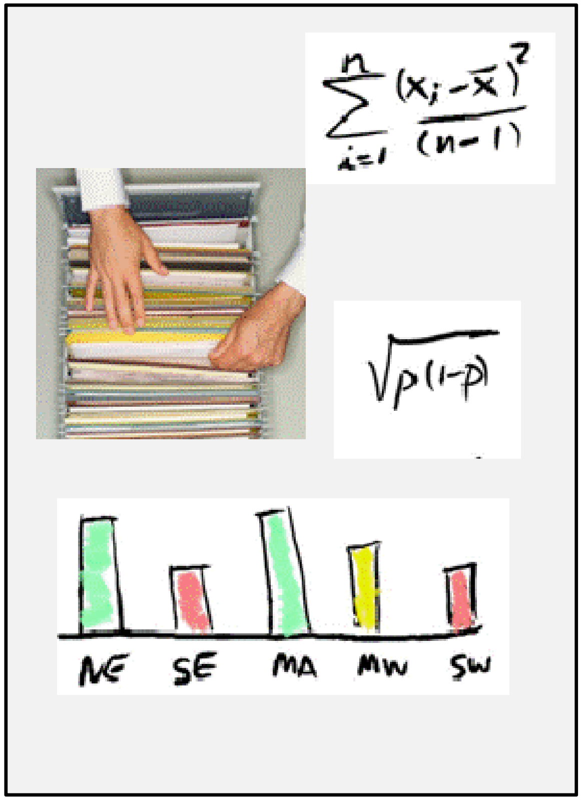
Problem Dimension	Traditional Applied Statistics	Statistical Engineering
Criticality to the Organization	Low - Medium	High
Impact – Financial, Process Performance, Customer, Social, and Environmental	Low - Medium	High
Number of Departments, Groups, or Functions Involved	Few	Several
Complexity – Technical, Political	Low	High
Sources of Information and Data	Few	Many
Number of Tools Involved	Few	Many
Use of Information Technology	Some	Essential
Need for Sustainability	May be needed	Essential

Why Statistical Engineering?

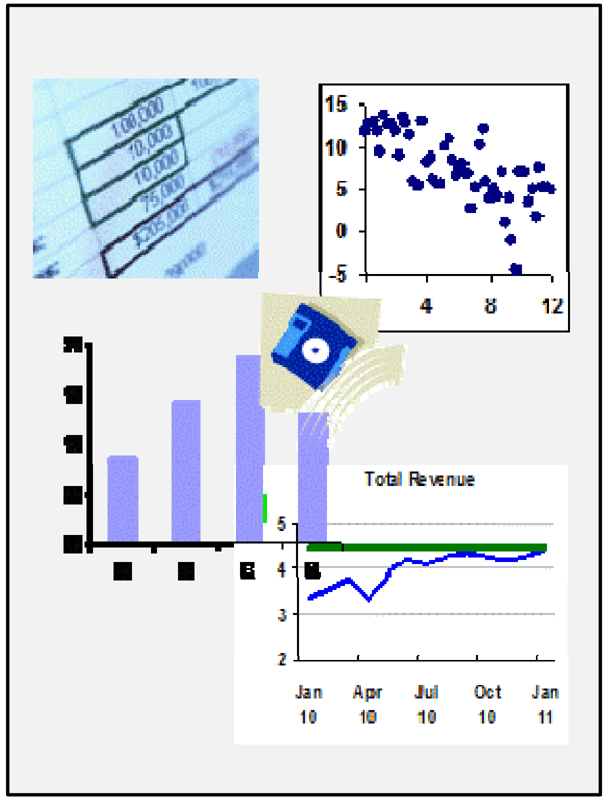
Mark Zabel

Proliferation and Accessibility of Data

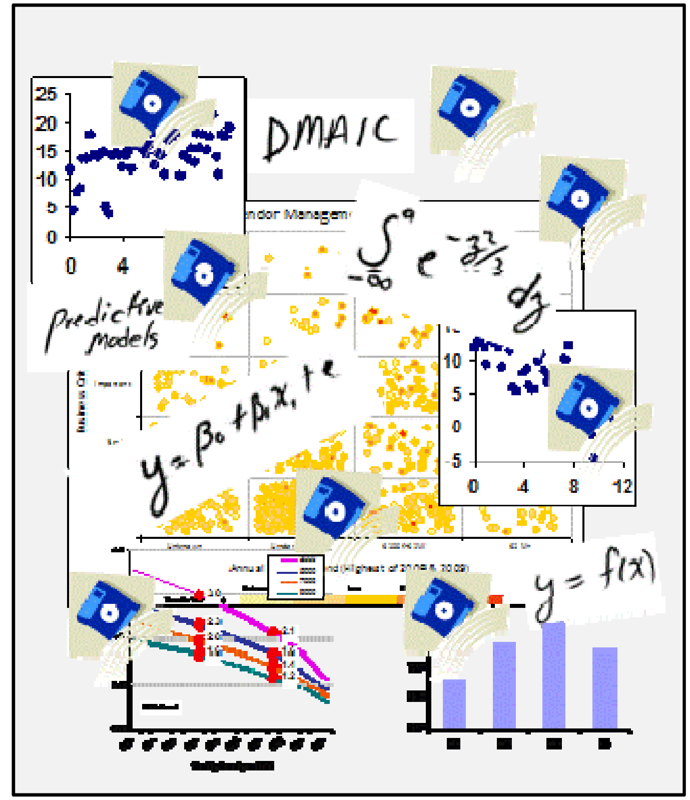
Yesterday



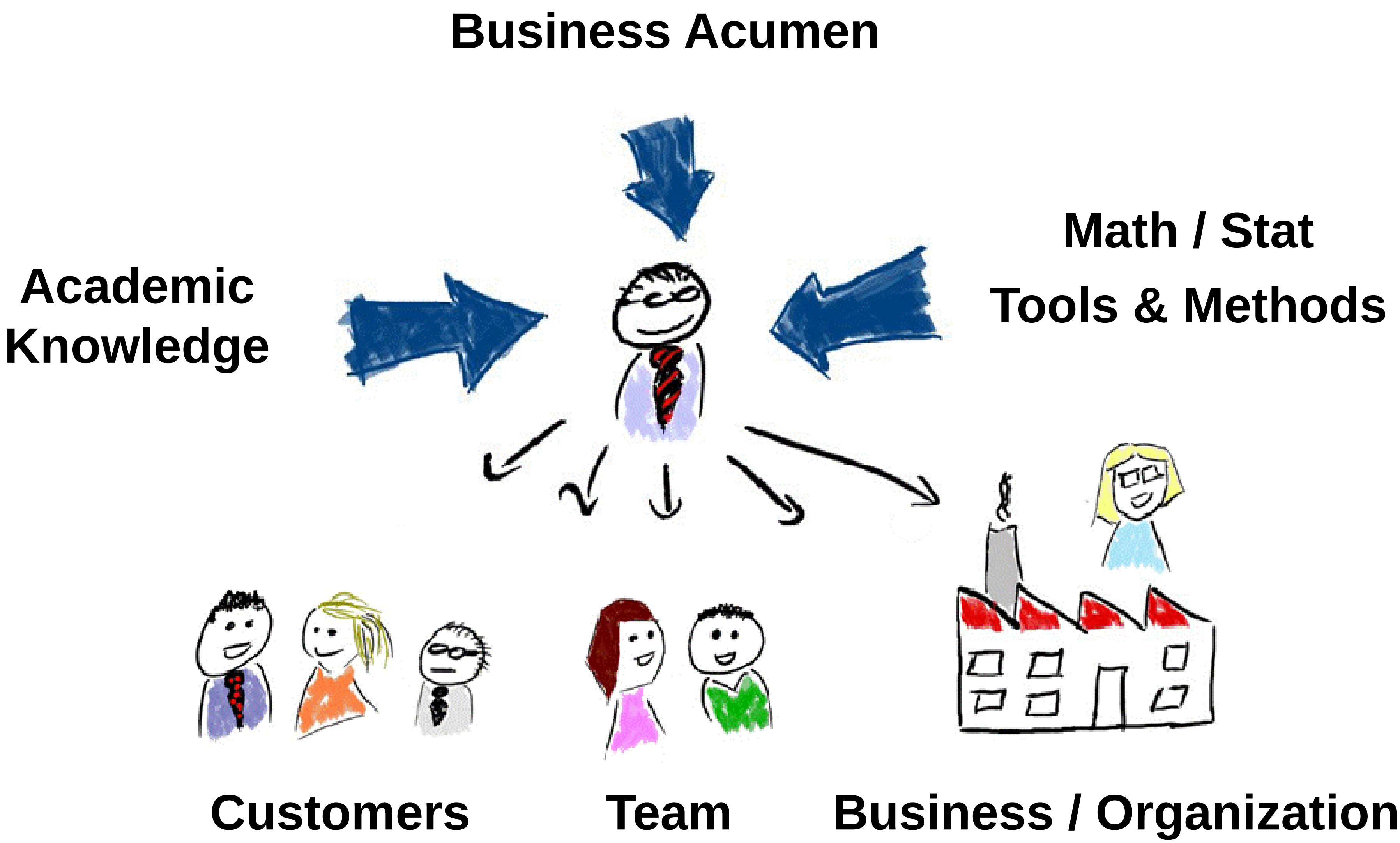
Today



Tomorrow

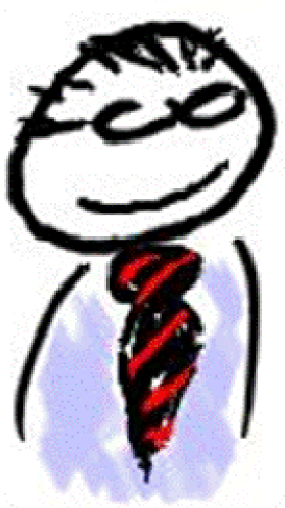


Statistical Engineers Connect



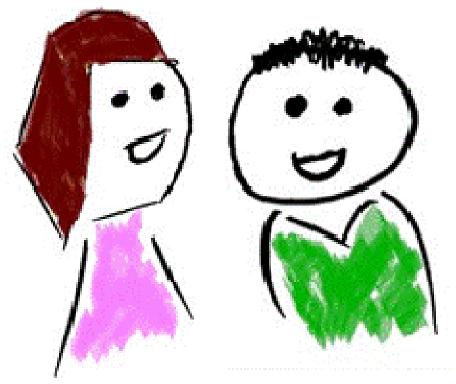
Benefits to the Individual Statistician

Individual Statistician



- **Greater collaboration and more acceptance of methods**
- **Deliver better results**
- **Experience drives new applications for academic work**
- **More opportunities for leadership positions**
- **Higher value to business & organization**

Team

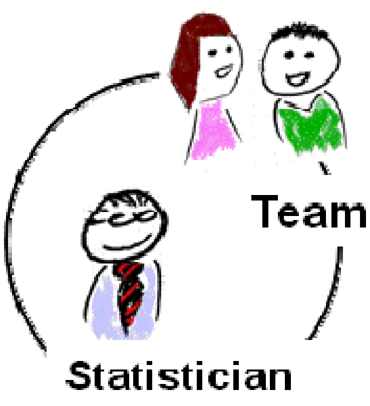


Statistical methods strengthen team delivery

- **Better business cases**
- **Better data and metrics**
- **More robust models**
- **Better known uncertainty and risk**
- **Efficient design and analysis**
- **Better reporting and other quantitative deliverables**

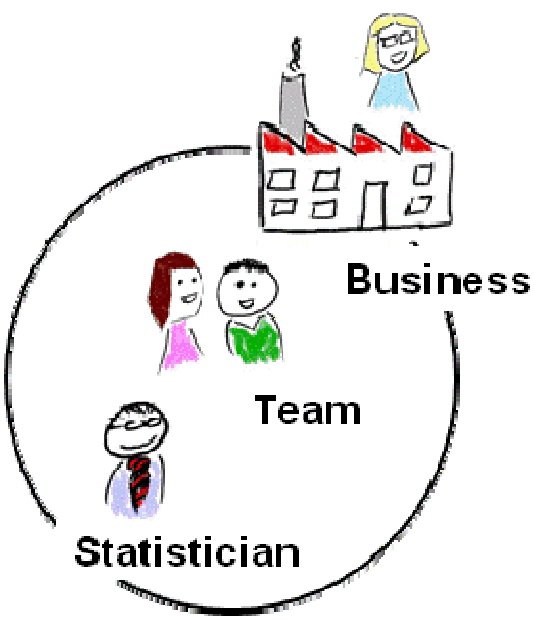
Benefits to the Business / Organization

Business

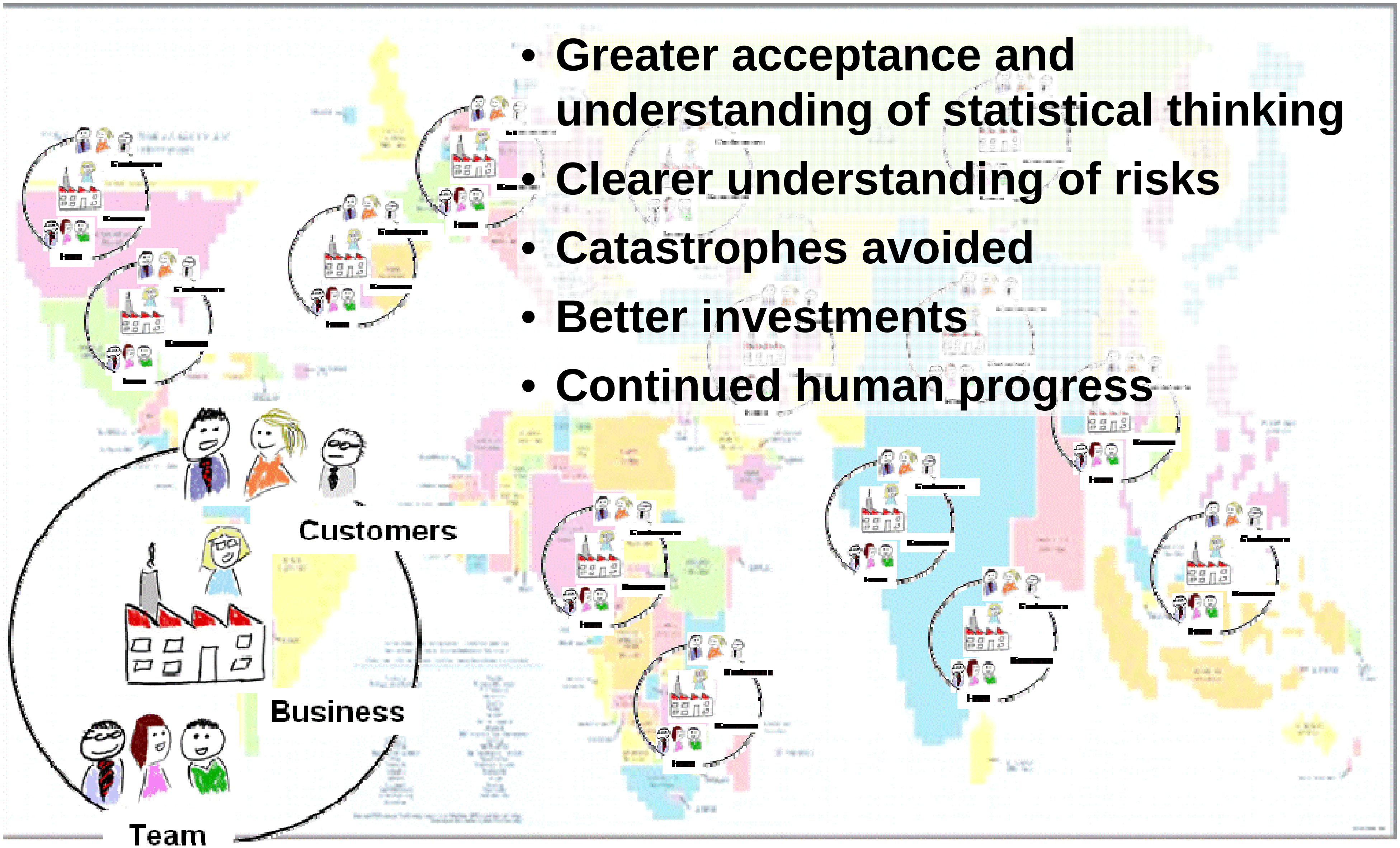


- **Better delivery = greater ROI**
- **Better and more informed decisions**
- **Lower uncertainty & risk**
- **Reduced costs**
- **Higher degree of fact-based engagement**
- **Better analytics**
- **Added insight into business and customer needs and wants**

Customers



- More robust, higher quality products & services
- Business partners who make better decisions
- Wants and needs are better known and better addressed
- Clearer insight
- Better decisions



- **Greater acceptance and understanding of statistical thinking**
- **Clearer understanding of risks**
- **Catastrophes avoided**
- **Better investments**
- **Continued human progress**

The How of Statistical Engineering?

Geoff Vining

- Agency
  - Past: Mechanical – Aerospace Driven
    - "Deterministic" Universe
    - Statistics not Always Appreciated
  - Current: Need to Quantify Uncertainty
    - Major Driver: Risk Management
    - Agency Requirements:
      - Rigorous
      - Scientifically defendable
    - Provides an Excellent Opening for Statistical Engineering

- Engineering
  - Example: To Estimate a Straight Line, Why Do We Use Ten or More Un-Replicated Levels?
    - "Replicates introduce error" (actually, they provide estimate of true error)
    - Only two levels needed to estimate the line; three levels allow estimate of lack-of-fit
  - Need for Statistical Thinking
    - All work occurs in systems of interconnected processes
    - Variation exists in all processes
    - Keys to success: understanding and reducing variation

- Statisticians
  - Passive/Active
  - Periphery/Central
  - Purely Data Analysis/Problem Solver
    - Problem Definition/Clarity
    - Systems Thinking
    - Tactical Deployment of “Statistical Tools”
  - Non-Essential Team Member/Full Colleague
  - Perception: Part of the Problem/Part of the Solution

- Complex Problems Require Appropriate Solutions
  - Team Approach to Solutions
  - Clear and Precise Problem Definitions
  - Systems Thinking
  - Understanding Sources of Variation
  - Appropriate Data
  - Tactical Deployment of Analytics
- Statistical Engineering Is an Appropriate Approach!

BUILDING CAPABILTY

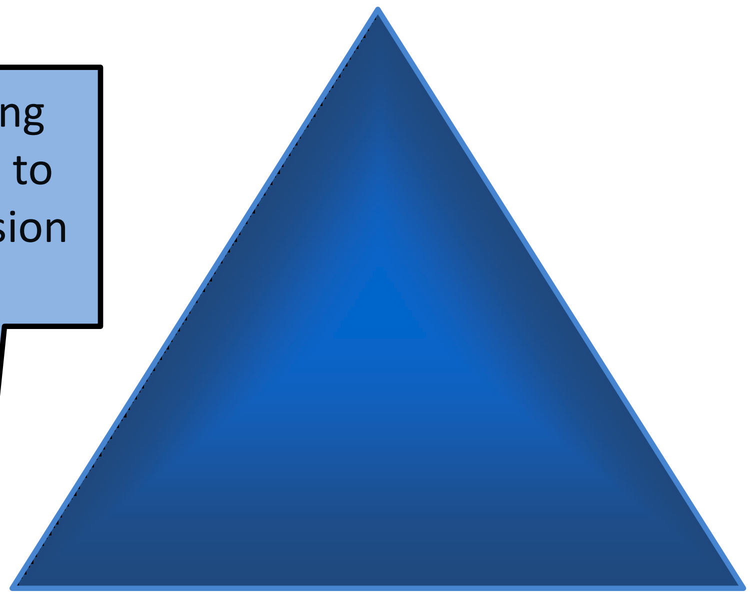
Adding Value

Mark Hutchinson

Capabilty Strategy

- Seek key leadership roles:
  - Agency, Center, Program/Projects
  - Professional Organizations

LEADERSHIP



Provide key engineering expertise and services to projects to ensure mission success.

- Participate in
  - Early Innovation
  - Game Changing & Cross Cutting Technologies
- Partnerships
  - Academia
  - Industry

ENGINEERING

TECHNOLOGY

Deliberate Excellence

- Overarching strategy for technical excellence and quality that is being worked regularly despite the day-to-day activities
- Establish Organizational long term strategy as a growth in technical capability
- Invest in People
  - Most Valuable Resource
  - Build Technical Expert/Technical leadership
- Invest in Tools/Procedures/Practice

NASA Langley Research Center

Director
Lesa Roe

Deputy Director
Steve Jurczyk

Chief of Staff
Cynthia C. Lee

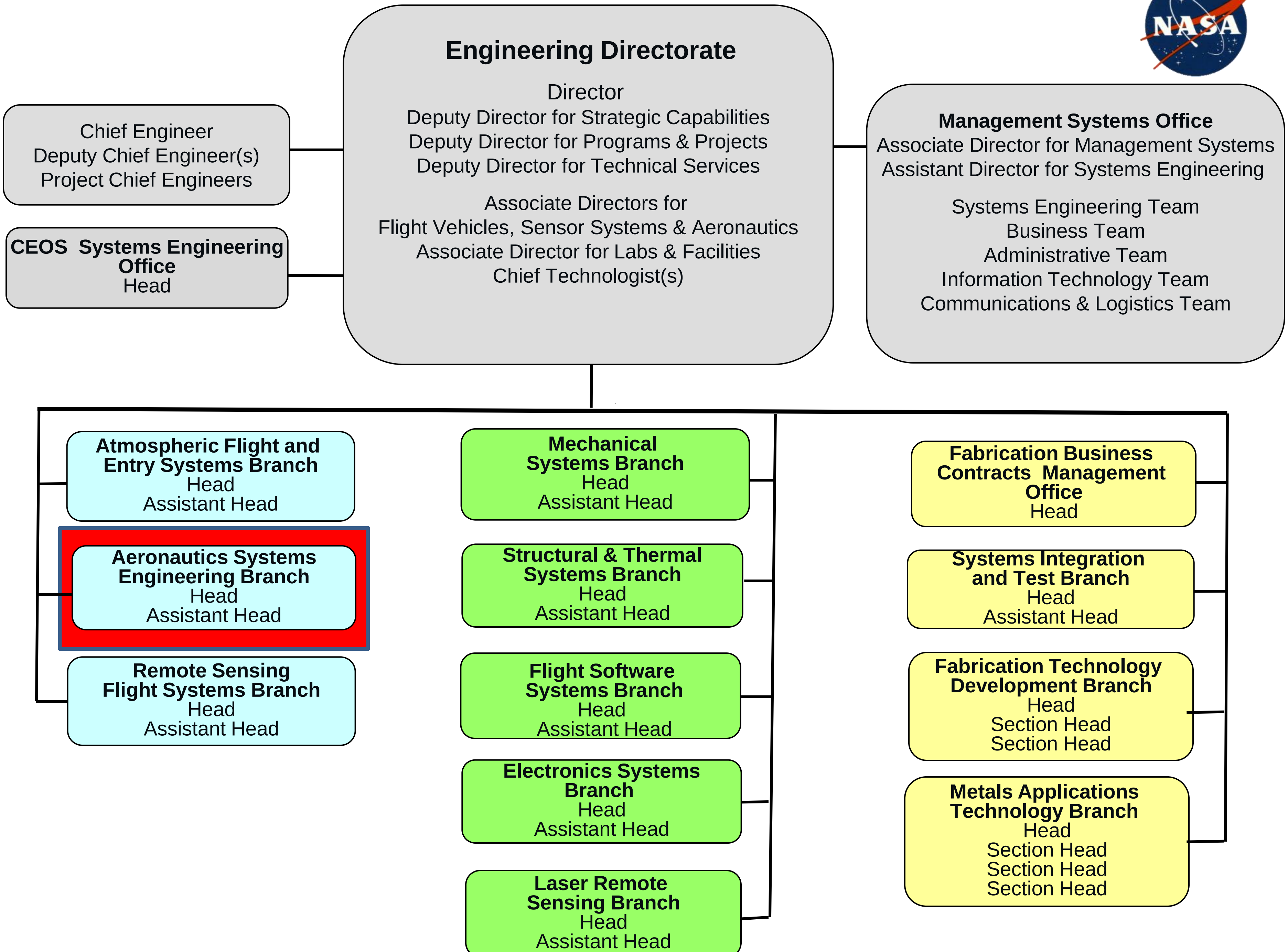
MISSION

- Aeronautics Research
- Science
- Exploration and Space Operations
- Flight Projects
- Systems Analysis and Concepts
- Strategic Relationships
- Research Services
- Ground Facilities and Testing
- Engineering
- Research
- Center Operations
  George B. Finelli

MISSION SUPPORT

- Safety and Mission Assurance
- Systems Management
- Procurement
- Chief Counsel
- Equal Opportunity
- Human Capital Management
- Strategic Communication
- Chief Financial Officer
- Chief Information Officer

The Langley Research Center (LaRC) pioneers the future in space exploration, scientific discovery, and aeronautics through research and development of technology, scientific instruments and investigations, and exploration systems.


Engineering Directorate
Director
Deputy Director for Strategic Capabilities
Deputy Director for Programs & Projects
Deputy Director for Technical Services
Associate Directors for
Flight Vehicles, Sensor Systems & Aeronautics
Associate Director for Labs & Facilities
Chief Technologist(s)
Chief Engineer
Deputy Chief Engineer(s)
Project Chief Engineers
CEOS Systems Engineering Office
Head
Management Systems Office
Associate Director for Management Systems
Assistant Director for Systems Engineering
Systems Engineering Team
Business Team
Administrative Team
Information Technology Team
Communications & Logistics Team
Atmospheric Flight and Entry Systems Branch
Head
Assistant Head
Aeronautics Systems Engineering Branch
Head
Assistant Head
Remote Sensing Flight Systems Branch
Head
Assistant Head
Mechanical Systems Branch
Head
Assistant Head
Structural & Thermal Systems Branch
Head
Assistant Head
Flight Software Systems Branch
Head
Assistant Head
Electronics Systems Branch
Head
Assistant Head
Laser Remote Sensing Branch
Head
Assistant Head
Fabrication Business Contracts Management Office
Head
Systems Integration and Test Branch
Head
Assistant Head
Fabrication Technology Development Branch
Head
Section Head
Section Head
Metals Applications Technology Branch
Head
Section Head
Section Head
Section Head

ASEB's Functional Statement

The *Aeronautics Systems Engineering Branch* provides a lead role in providing experimental hardware, advanced sensors and measurement systems using a systems engineering approach that enable our customers to *gain knowledge* by simulation of aerospace concepts for aerodynamic and structures research.

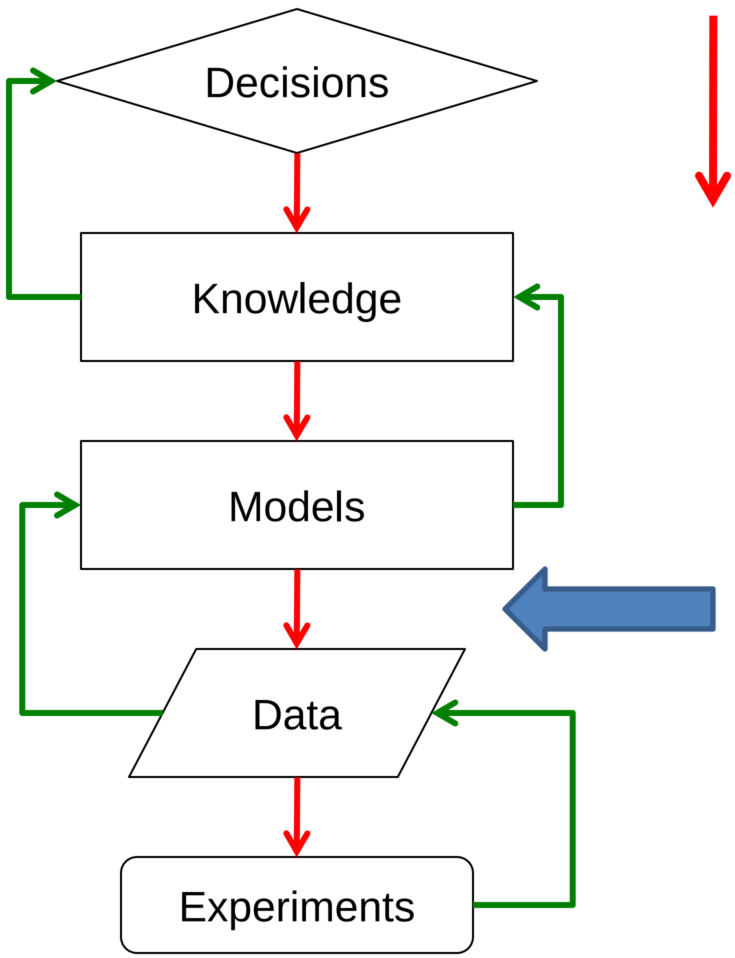
Investment in People

- **Encouragement of**
  - Innovation
  - Education
  - Professional Society Participation
  - Affiliation with Academia
  - Mentorship with Students
- **Publish**
- **Difficult Technical Challenges**
  - Is it a big enough problem?

80/20 Rule

- 20% of employee's time dedicated to capability growth
  - Training
  - Publishing
  - Mentoring
- 80% to Product Delivery

Planning and Execution Model



Requirements Flow Down

Things need to <u>Know</u>

How well we need to know?

Normal Request

Project Execution

Things need to <u>Do</u>

Heilmeier Questions

Excerpt from IEEE Spectrum Article, June 1994

- **What are you trying to do?**
  - **Articulate your objectives using absolutely no jargon**
- **How is it done today, and what are the limitations of current practice?**
- **What's new in your approach, and why do you think it will be successful?**
- **Who cares?**
  - **If you are successful, what difference will it make?**
- **What are the risks and the payoffs?**
- **How much will it cost? How long will it take?**
- **What are the midterm and final "exams" to check for success?**

Statistical Engineering Questions

Program and Project Definition

- What are the precise objectives?
  - Are the objectives quantifiable, detectable, measurable?
  - What are we seeking to learn, new knowledge sought?
  - How will we know when we have learned it?

Technical Risk Management

- How well do we need to know the answer(s) (precision)?
  - What risk are we willing to accept if we are wrong about or conclusions?
  - What are the consequences if we are wrong?

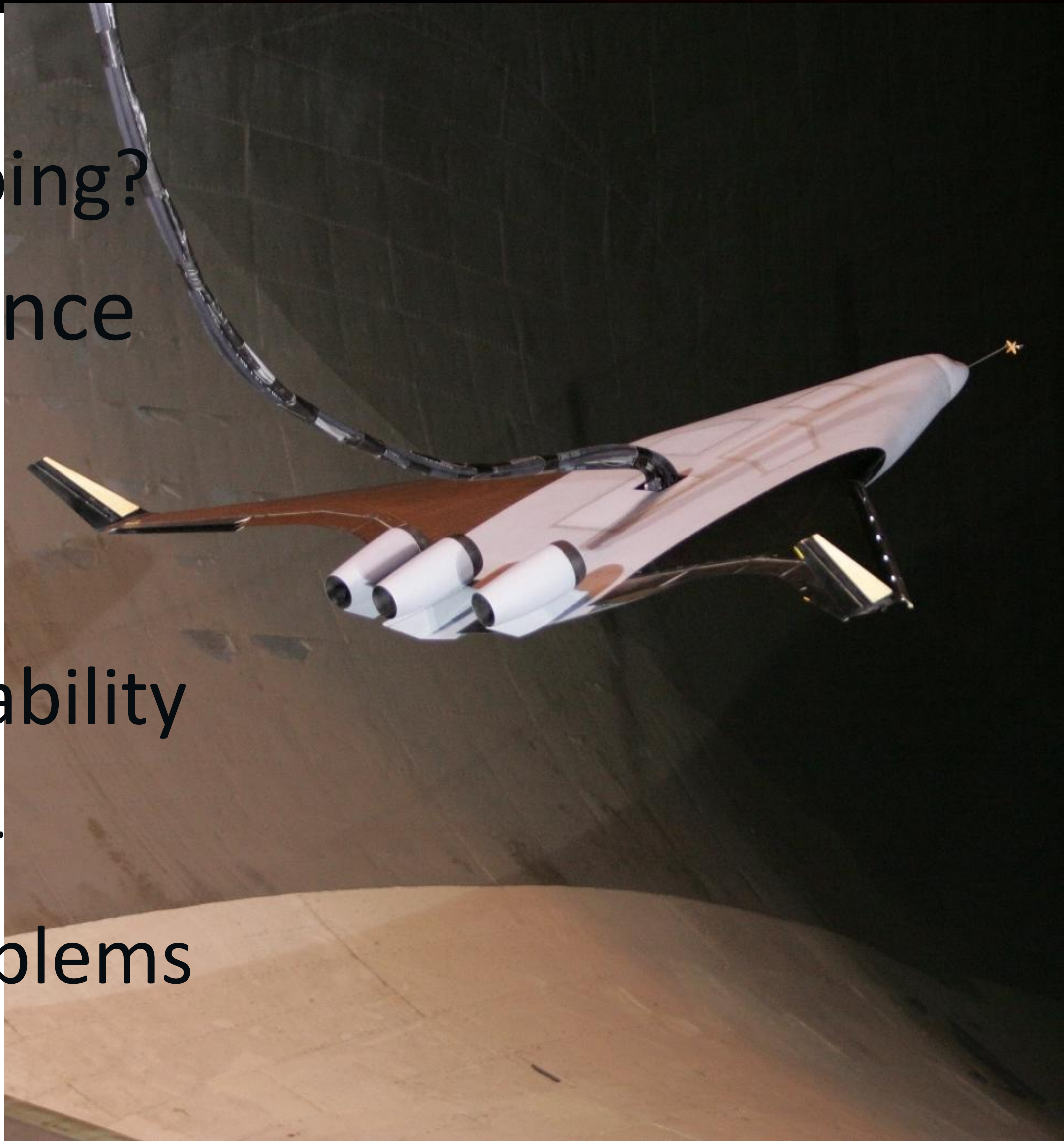
Planning and Execution

- Do the methods support rigorously link to the objectives and risk?
- Does the allocation of resources support the objectives and risk?
  - Are the resources justifiable and defendable?

Questions apply recursively in the vertical direction through systems and subsystems and horizontally throughout project phases

Building Capability

- Vision
  - Where are you going?
- Deliberate Excellence
  - Plan to Excellent
- People-
  - Most Valued Capability
- Challenging work-
  - Hard to solve problems



Infusing Statistical Engineering

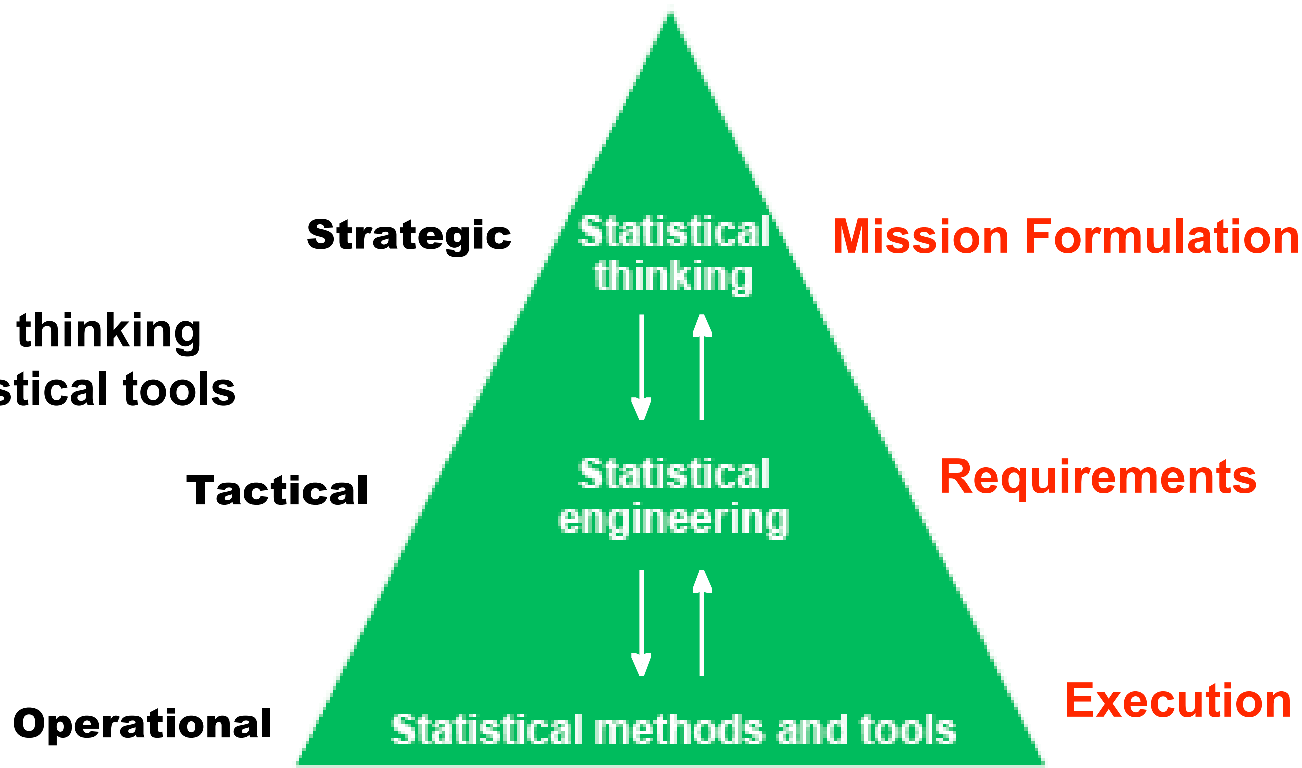
Peter A. Parker, Ph.D., P.E.
National Aeronautics and Space Administration
Langley Research Center Engineering Directorate
Aeronautics Systems Engineering Branch

May 4, 2011
NASA Statistical Engineering Symposium
Williamsburg, Virginia

Statistical Engineering at NASA

- Engineering discipline to efficiently gain knowledge through strategic resource investment
- Applies systems thinking to high-level, well-defined objectives
- Synergistic combination of existing tools to solve complex problems

Drives statistical thinking
and guides statistical tools



Parker (2008), "Infusing Statistical Engineering in Programs and Projects," *NASA LaRC White Paper*
Figure: Hoerl, R.W. and Snee, R.D. (2010) "Closing the Gap," *Quality Progress*

Defining Rocket Motor Requirements

- Leveraging heritage design for a new application and requirements

Original Question

- How to measure roll torque during a firing to size the reaction control system?
- Impacts requirements, cost (lifecycle), schedule, volume, mass (payload)

Statistical Engineering Applied

  - System level – studied apparatus, available theory and data
  - Integrated design of experiments with force measurement expertise
  - Measured roll torque, quantified uncertainty, reduced design risk
- Internally generated ballistics could not be rigorously isolated
  - Modifications to the firing duty cycle fully achieved objectives
- Embedded in processes, software, and training

Gentz (NESC), Roberts (NESC), Parker (LaRC), Rhew (LaRC), Jones (MSFC), McLennan (ATK)

Motivation for Statistical Engineering

- Consistent methodological framework for research and development
  - teachable, repeatable, scalable - not idiosyncratic
- Benefits of successful implementation
  - Improved Decision-making - risk-informed and defendable
  - Technical excellence – unequivocally define objectives, integrity of results/knowledge obtained
  - Organizational excellence - strategically applying resources
- Opportunity for statisticians to make greater contributions in achieving strategic organizational objectives

Changing the Role of the Statistician

Required for Effective Statistical Engineering

Project Lead \ Statistician	Passive	Active
Active	Statistician as **Helper**	Statistician as **Colleague**
Passive		Statistician as **Leader**

Figure: Hunter, W.G. (1981) “The Practice of Statistics: The Real World Is an Idea Whose Time Has Come,” *The American Statistician*

Starts with Fundamental Questions

- Heilmeier used as a preflight checklist for successfully launching a research project to curb and clarify both the enthusiasm of the researchers and to evaluate the resource demands of the project managers

- **What are the precise objectives?**
  - **What are we seeking to learn?**
  - **Are the objectives quantifiable, detectable, measurable?**
  - **What is the impact if you are successful?**

- **How well do we need to know the answers?**
  - **How much risk are we willing to accept in being wrong?**
  - **What are the consequences if we are wrong?**

- **Do the methods rigorously link to the stated objectives and risk?**
  - **Are the resources justifiable and defendable?**

Questions apply recursively in the vertical direction down to systems and subsystems and horizontally throughout project life-cycle

Shapiro (1994) "George H. Heilmeier," *IEEE Spectrum*

Using NASA Satellite Data and Models for Socioeconomic Benefits

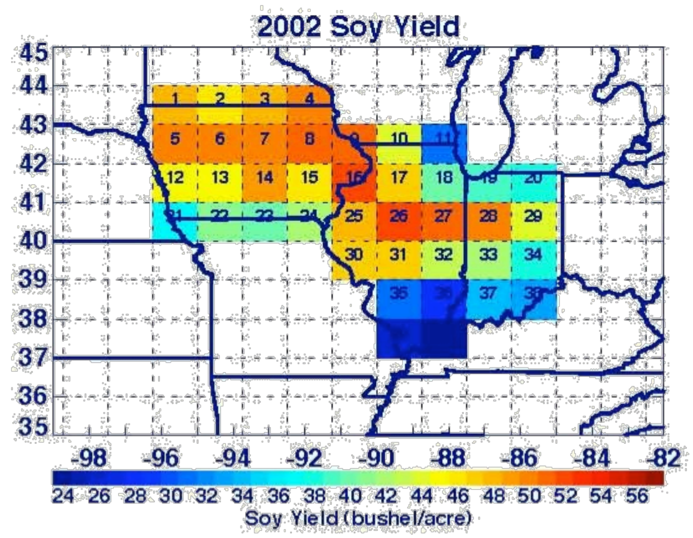
- **Research into using ozone satellite measurements (Tropospheric Ozone Residual) to improve soybean crop yield**

Original Question

- **Can we develop a correlation between satellite measurements and ground-based measurements of ozone?**

Statistical Engineering Applied

- **Synthesized multidisciplinary team member ideas on the objectives to clearly define the research questions and approach**
- **Satellite and ground provide different types of useful information**
- **Modeled yield as a function of temperature, soil moisture, and ozone (satellite and ground)**
- **Impact: Provided new framework, impetus for additional research**



Fishman (LaRC), Parker (LaRC), Vining (VT), Ainsworth (USDA), et al.

Vital Implementation Elements

- **Leadership**
  - **Requires leadership to convert a "good idea" into "the new way we do business"**
  - **Articulate motivation, communicate expectations, be accountable**
- **Core Competency**
  - **Discipline experts matrixed across programs/projects**
  - **Multidisciplinary skills are required**
  - **Teaming and communication skills are critical**
- **Equipping People with Knowledge and Tools**
  - **Broaden awareness of this discipline**
  - **Consult with researchers and lead engineers, equip practitioners**
- **To be seen as value-added and measure effectiveness**

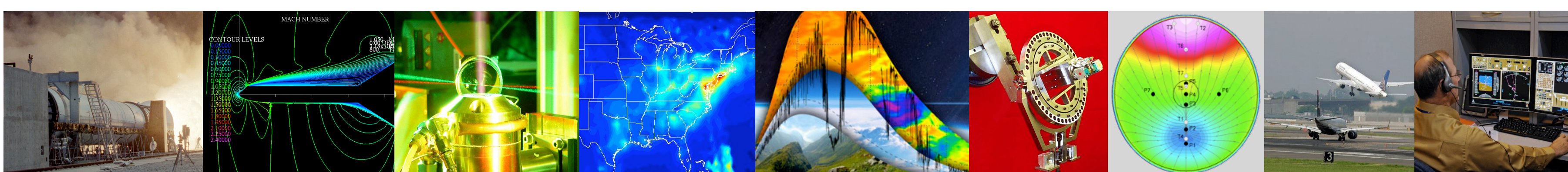
Hutto, G. (2005*), "White Paper 53rd Wing DOE Change Initiative, " and personal communications*
Kotter, J.P. (1996), "*Leading Change*," Harvard Business Press

Our Progress at NASA Langley

- **Building a statistical engineering capability takes a deliberate strategy**
- **Obtained leadership support through demonstrated benefits**
- **Growing a core team with multi-disciplinary competence**
- **Broadened knowledge of the discipline by recognized project impact**

<u>Areas we need to improve</u>

- **Inextricably link statistical engineering to organizational objectives**
- **Assist leadership to further engage and commit with specific actions**



Our Vision for Statistical Engineering

- Tactical Discipline that
  - Drives critical, statistical thinking at the strategic level
  - Guides statistical methods and research at the operational level
- Improves our effectiveness in accomplishing our mission
  - Defines the right questions
  - Guides strategic resource investment
  - Accelerates research and development
- Not a replacement for good science and engineering
- To promote best practices within our Agency and profession

Topics in Response Surface Model Adequacy *Assurance* and *Assessment*

Richard DeLoach
NASA Langley Research Center

The NASA Statistical Engineering Symposium
Williamsburg, VA
May 3-5, 2011

An Alternative Concept of *Quality* in Experimental Aeronautics

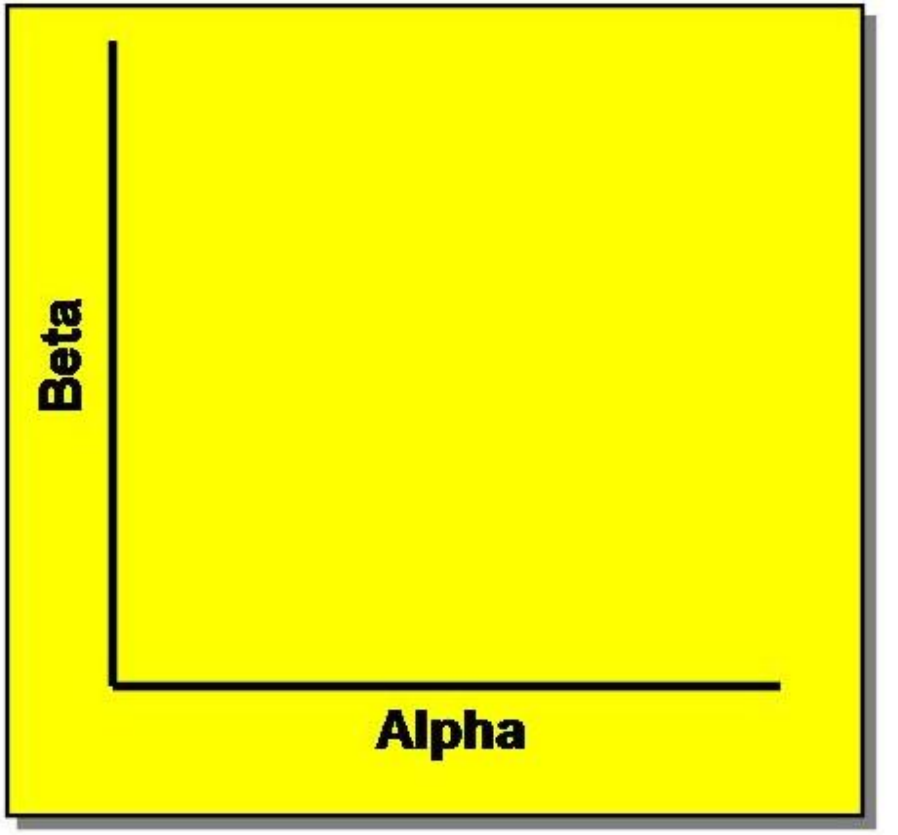
- **Traditional concept of quality in wind tunnel testing**
  - Data-centric: "Quality" means "Data Quality" in traditional testing
  - Associated with low levels of unexplained variance in a data sample
- **An alternative concept of quality**
  - Introduced to the Langley experimental aeronautics community in the mid-90's as the Modern Design of Experiments (MDOE)
  - Associated with inference error probability
  - "Quality" means "getting the right answer"
    - Low probability of inference error
    - Independent of quality of the data

Response Surface Modeling

- **Response Surface Models are mathematical functions representing responses (forces/moments, etc.) as a function of independent variables (AoA, Mach No., etc.)**
- **Quality is cast in terms of modeling adequacy**
  - **For an adequate model, no more than a specified percentage of response predictions are outside acceptable tolerance limits**
  - **Quality, or model adequacy, must be both *assured* and *assessed***
- **Model adequacy is *assured* through the design by**
  - **Data volume specification (*How many* points)**
  - **Site selection within the design space (*Which* points)**
  - **Number and selection of points to be replicated**
  - **Order in which the points are acquired**
- **Model adequacy is *assessed* by examination of residuals**

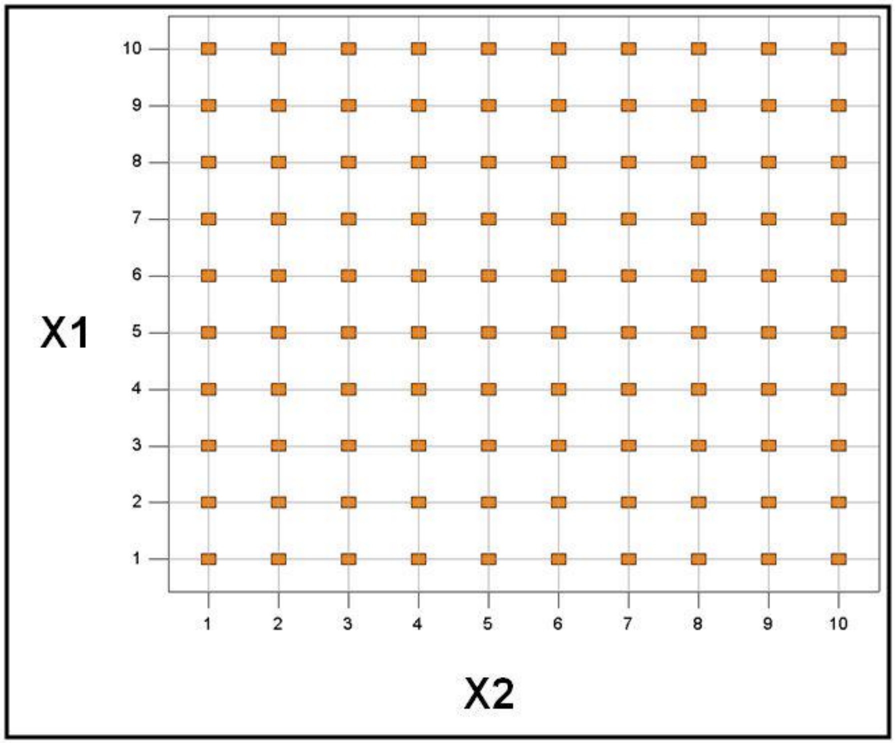
An Inference Space

- A Coordinate System
- One axis for each variable
- Each point represents a unique combination of variable levels
- A response surface is constructed "over" an inference space

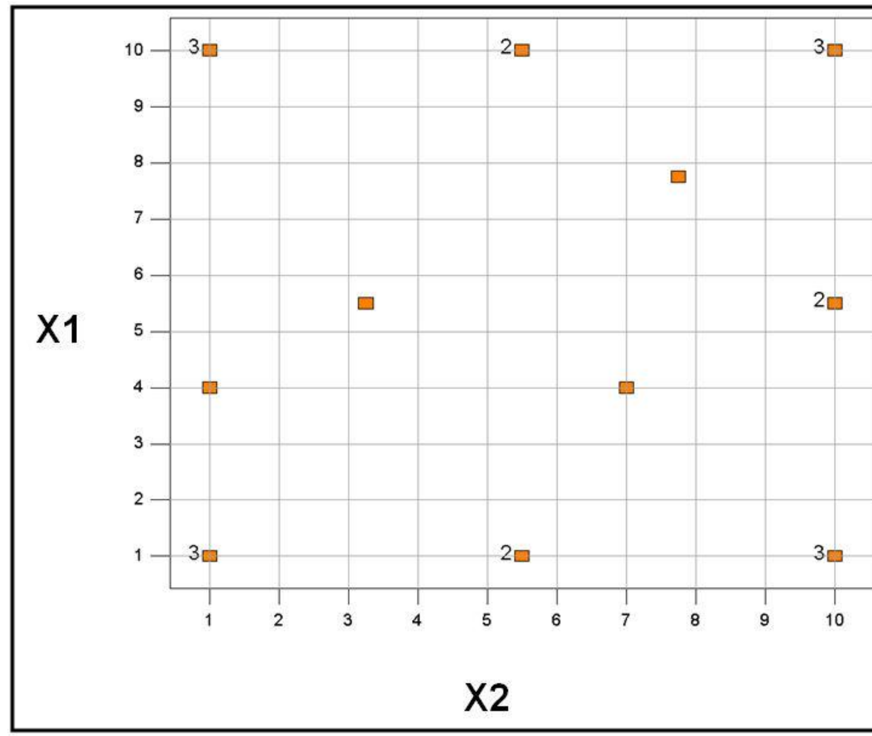


Design Space Comparisons

OFAT

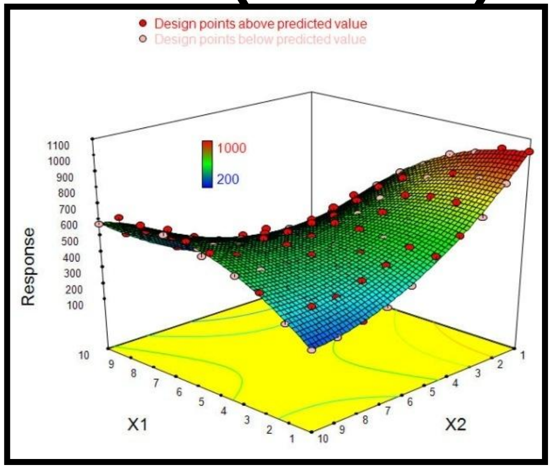


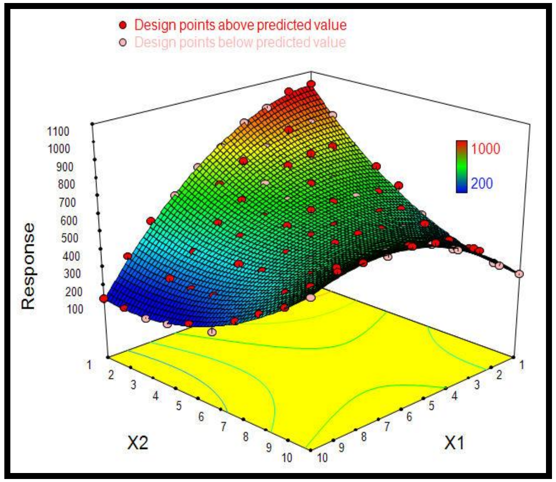
MDOE



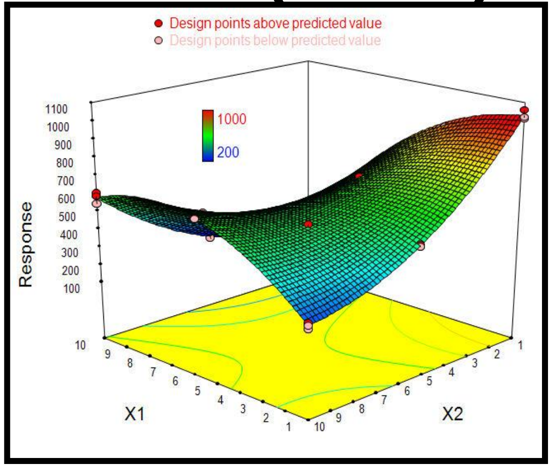
OFAT and MDOE Response Surfaces

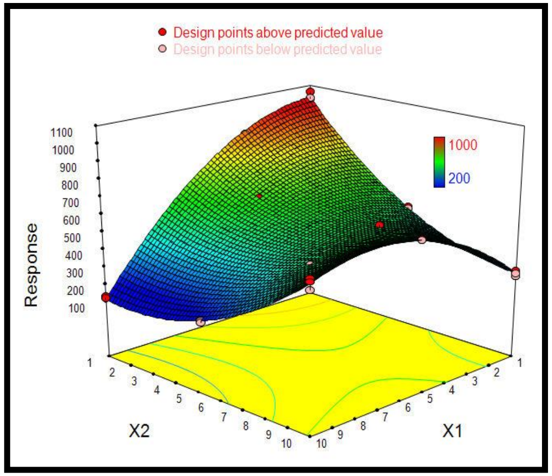
OFAT (100 Pts)





MDOE (22 Pts)



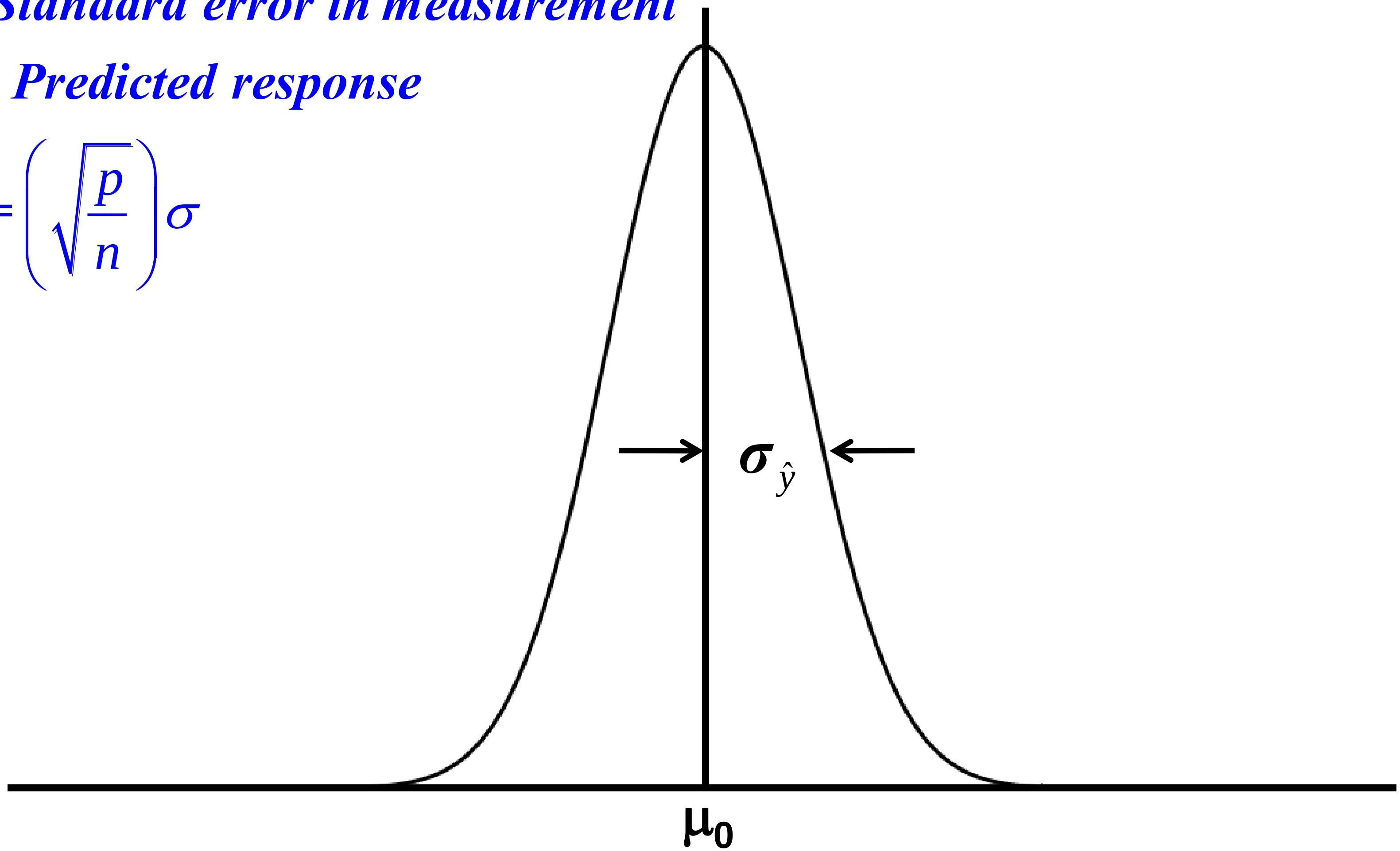


The Mathematics of Quality Assurance and Quality Assessment

σ : *Standard error in measurement*

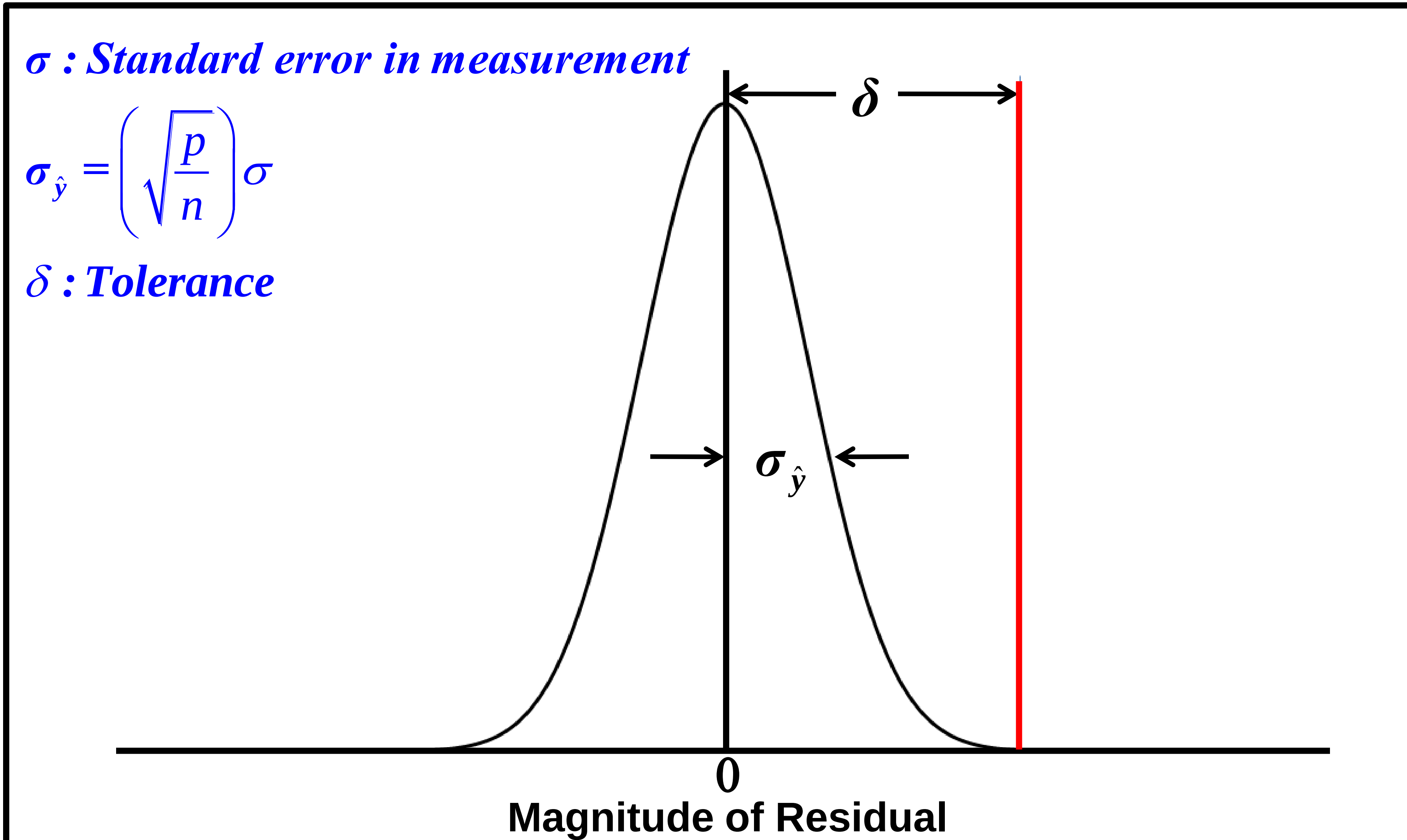
μ_0 : *Predicted response*

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$



Reference Distribution Under H_0

H_0: Null hypothesis that there is no difference between predicted and measured response (Residual is 0)



Reference Distributions for Residuals

Black – H_0: True residual is zero

Red – H_A: True residual is borderline unacceptable

σ : Standard error in measurement

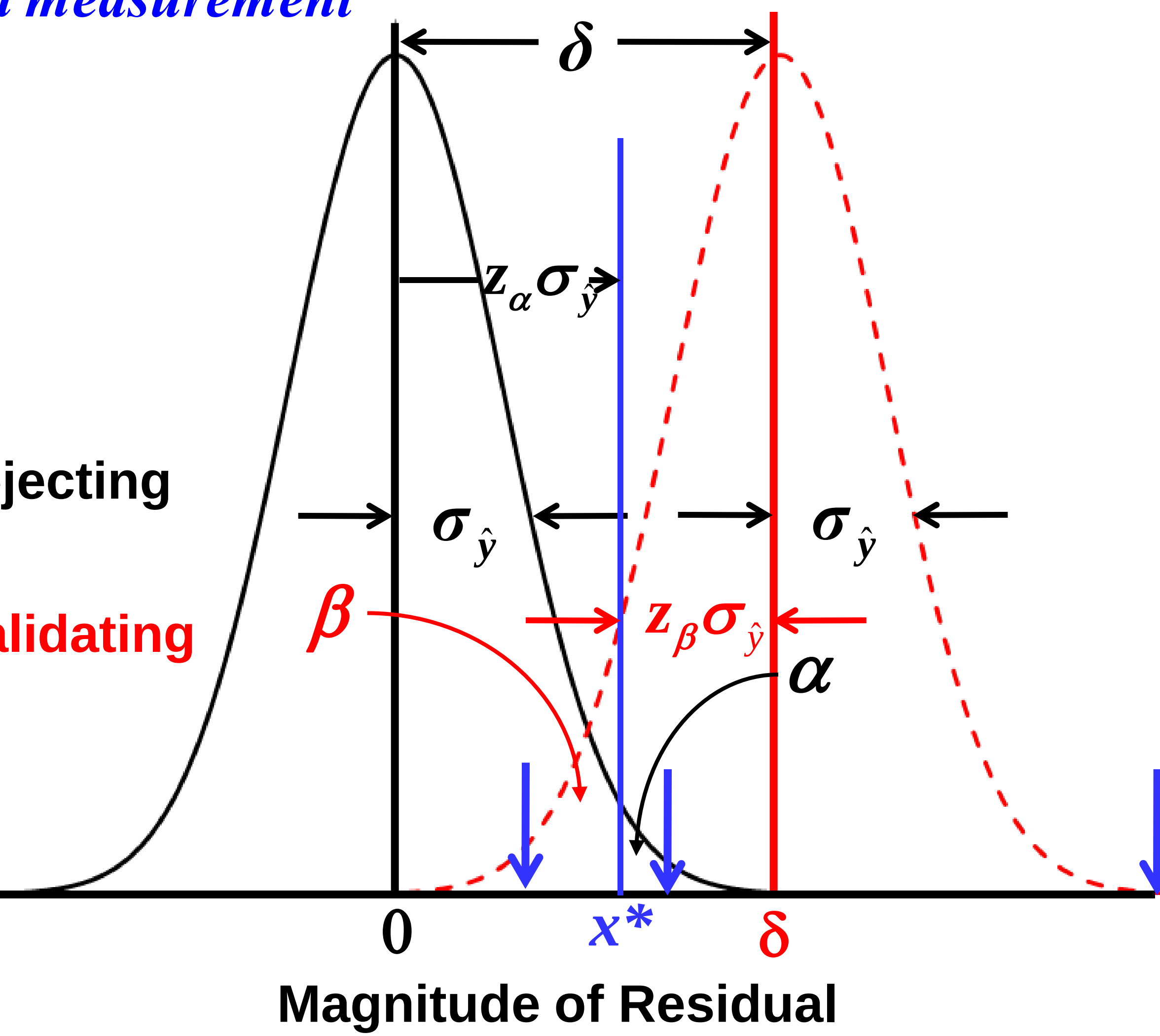
$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$

δ : Tolerance

x^* : Criterion

α is probability of rejecting adequate model

β is probability of validating erroneous model



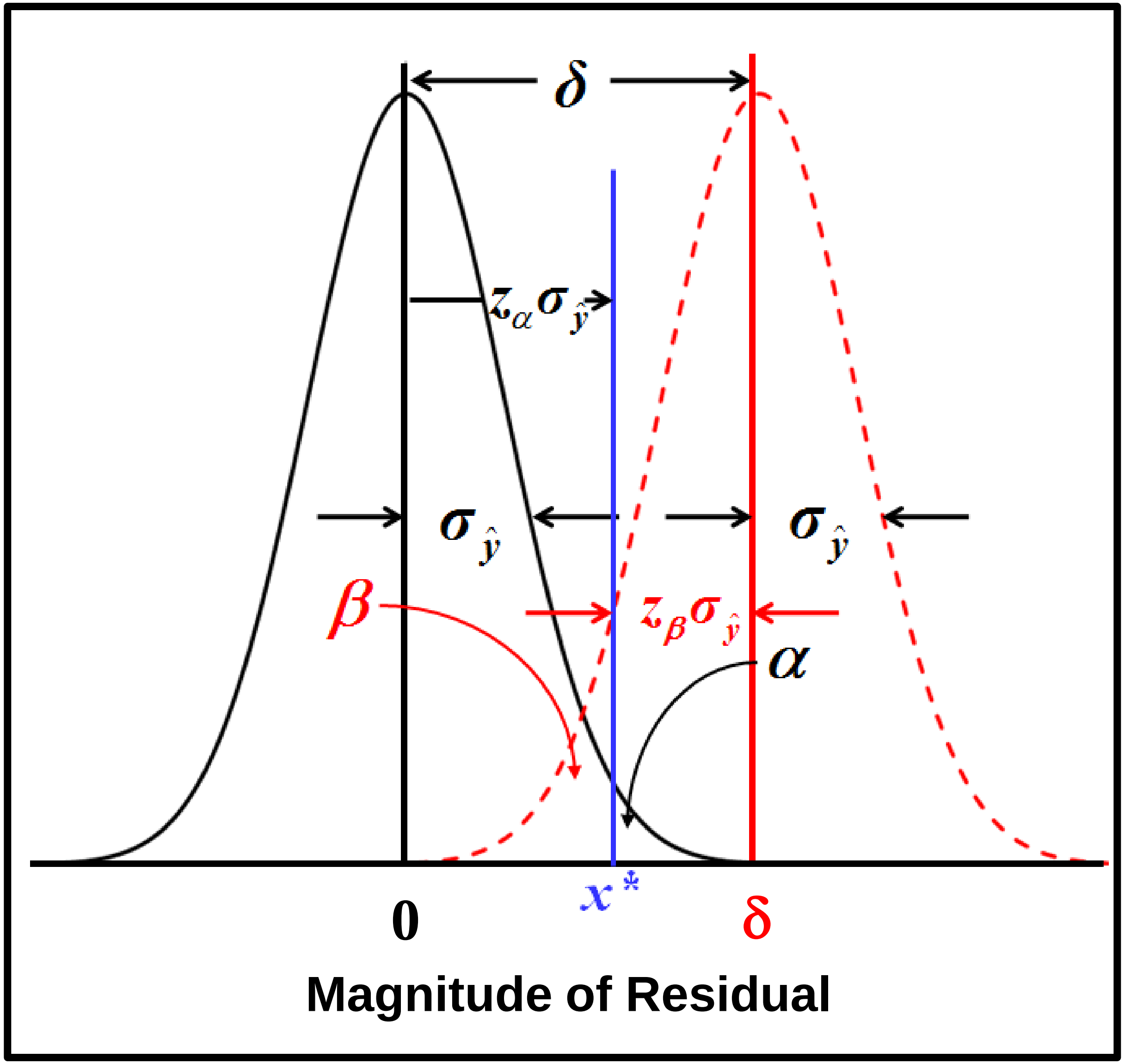
Data Volume Requirement

$$\delta = \left(z_\alpha + z_\beta\right)\sigma_{\hat{y}}$$

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$

$$\delta^2 = \frac{p\left(z_\alpha + z_\beta\right)^2 \sigma^2}{n}$$

$$\boxed{n = p\left[\left(z_\alpha + z_\beta\right)^2 \frac{\sigma^2}{\delta^2}\right]}$$



Data Volume Formula

Some Practical Difficulties

$$n = p\left[\left(z_\alpha + z_\beta\right)^2 \frac{\sigma^2}{\delta^2}\right]$$

- **The data volume formula depends on five quantities**
  - **Three (*p*, α, and β) can often be specified by the design consultant**
  - **The tolerance, δ, should be specified by the customer**
  - **The standard measurement error, σ, should be specified by the facility**
- **The customer often prefers to specify tolerance as a multiple of σ, rather than in absolute terms**
  - **A customer may feel comfortable saying his tolerance is "2σ"**
  - **He doesn't always feel he has to know what "σ" is to say this**

Incorporating Tolerance in Data Volume Estimates

- **Consider the general case, in which $\delta = K\sigma$, where K is a constant specified by the customer**
- **Note that a specific "*K*" may eventually evolve as an industry convention (about which more in a moment)**

$$\delta = K\sigma$$

$$n = p\left(z_\alpha + z_\beta\right)^2 \frac{\sigma^2}{\delta^2}$$

$$n = \left(\frac{z_\alpha + z_\beta}{K}\right)^2 p$$

Special Case for Tolerance, δ

- **We have for the general case in which $\delta = K\sigma$:**

$$n = \left(\frac{z_\alpha + z_\beta}{K} \right)^2 p$$

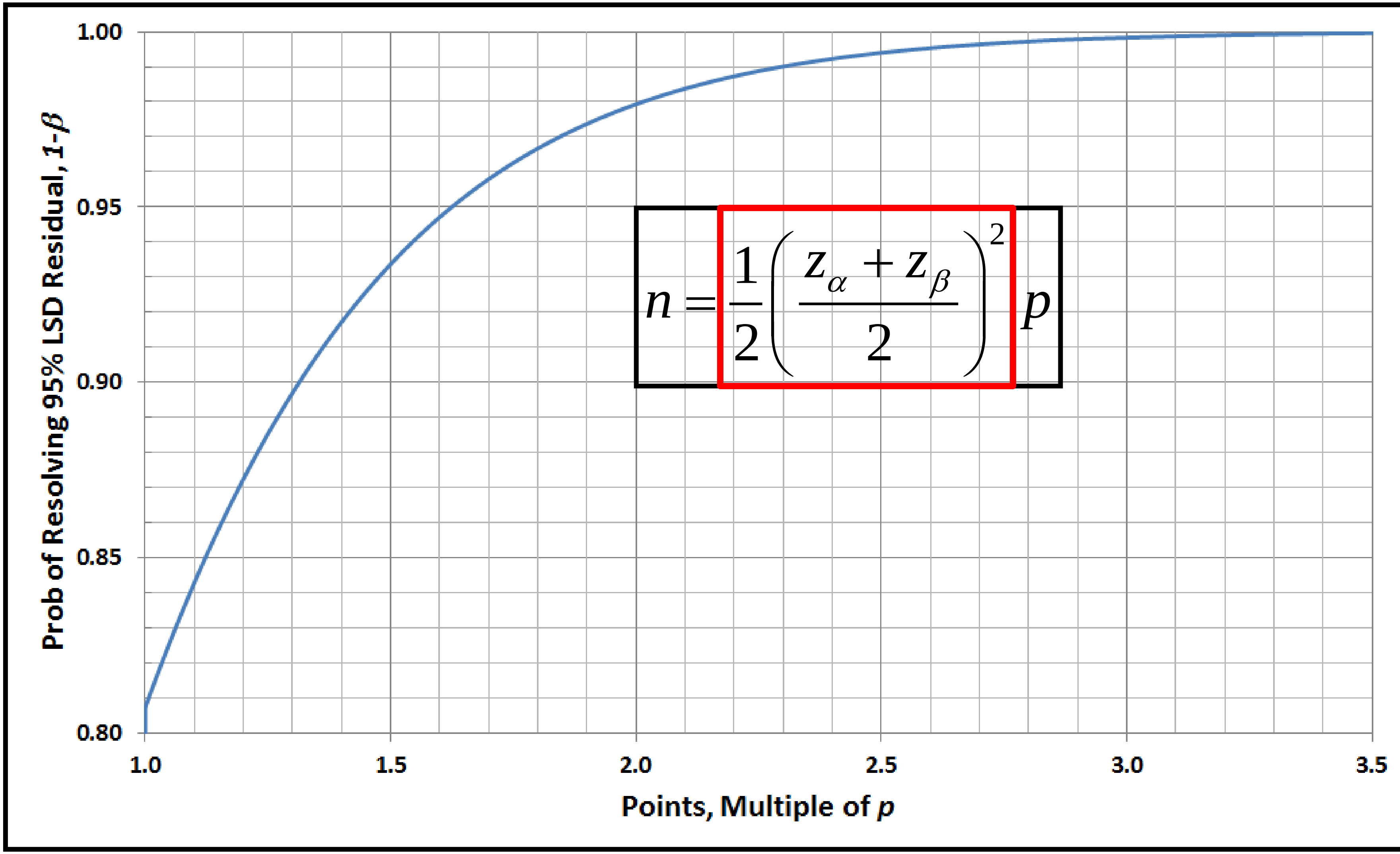
- **Let δ = 95% LSD (Least Significant Difference)**
  - **This is the smallest difference between two replicated measurements that can be resolved with 95% confidence**
  - **It may be regarded as a reasonable tolerance specification**

$$\delta = 95\%\ \mathrm{LSD} = 2\sqrt{2}\sigma \rightarrow K = 2\sqrt{2}$$

$$n = \frac{1}{2} \left(\frac{z_\alpha + z_\beta}{2} \right)^2 p$$

Model Term-Count Multiplier

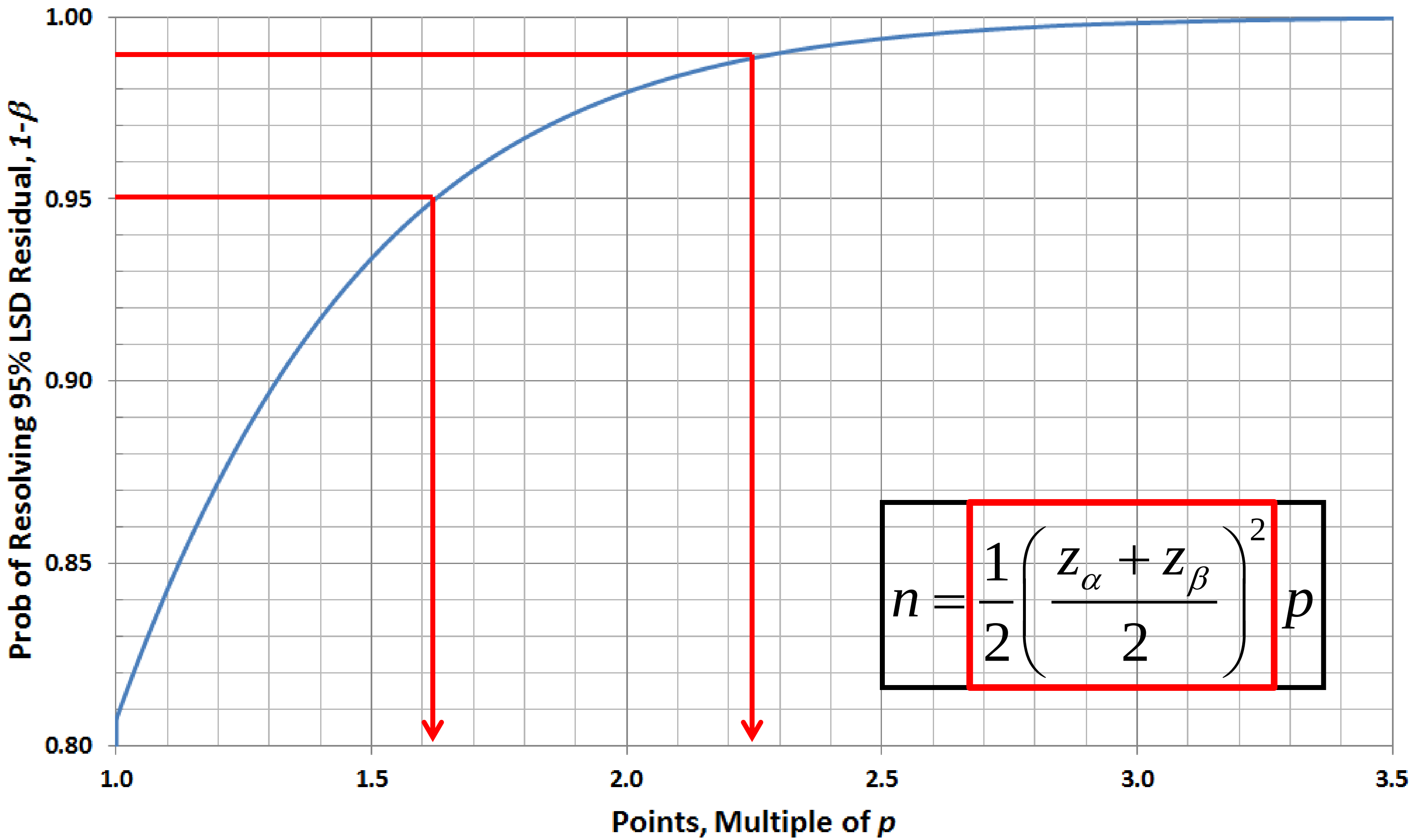
Minimum to Resolve 95% LSD with $\alpha = 0.05$



$$n = \boxed{\frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2} p$$

Model Term-Count Multiplier

Minimum to Resolve 95% LSD with $\alpha = 0.05$



$$n = \boxed{\frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2} p$$

Another Special Case for Tolerance, δ

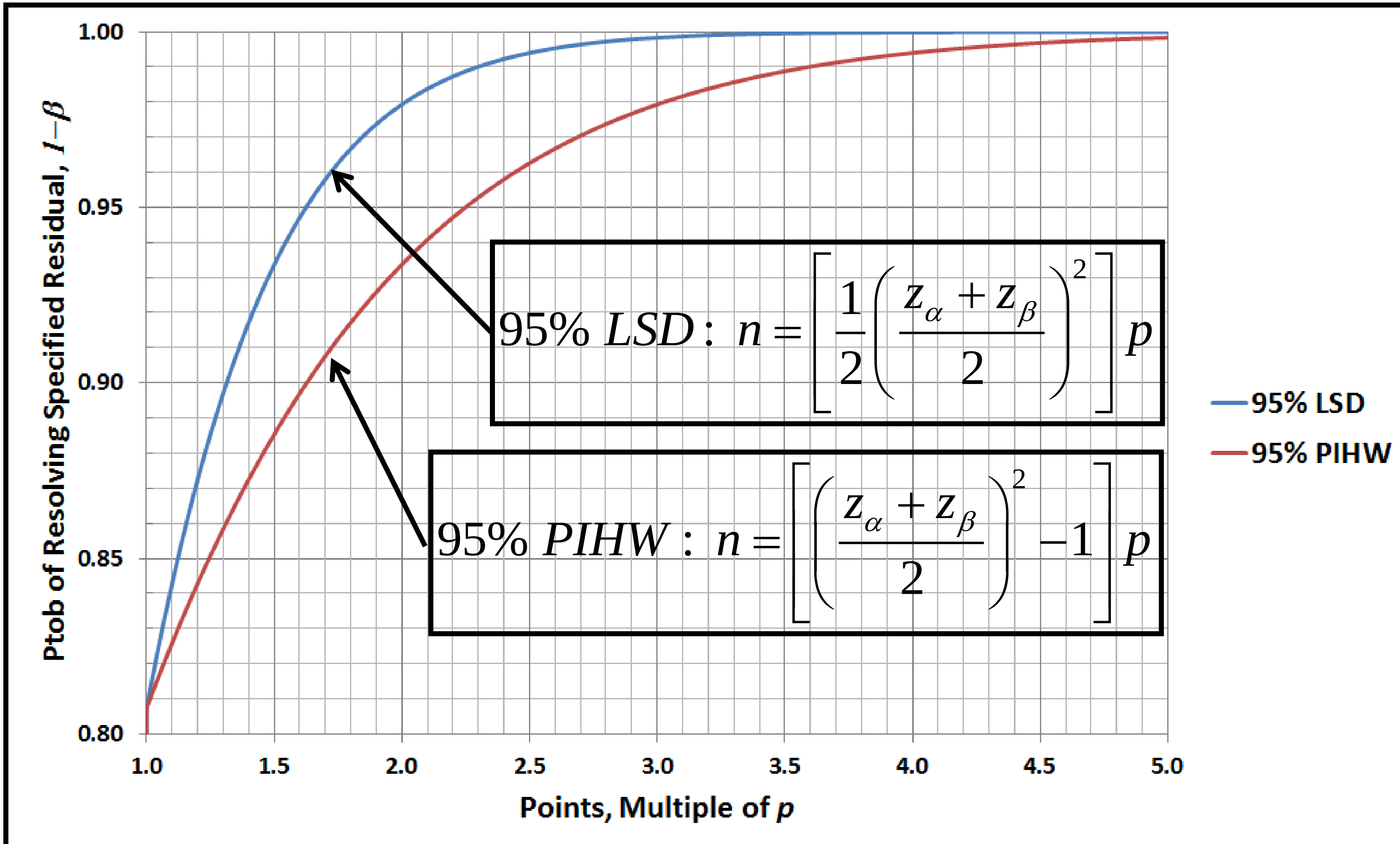
- **Let δ = 95% PIHW (Prediction Interval Half-Width)**
  - **This is the smallest difference between a physical measurement and a *model prediction* that can be resolved with 95% confidence**
  - **It is a convenient tolerance spec because most curve-fitting software packages compute this automatically**

$$\text{95\% PIHW} = 2\sqrt{\left(1+\frac{p}{n}\right)}\sigma \rightarrow K = 2\sqrt{\left(1+\frac{p}{n}\right)}$$

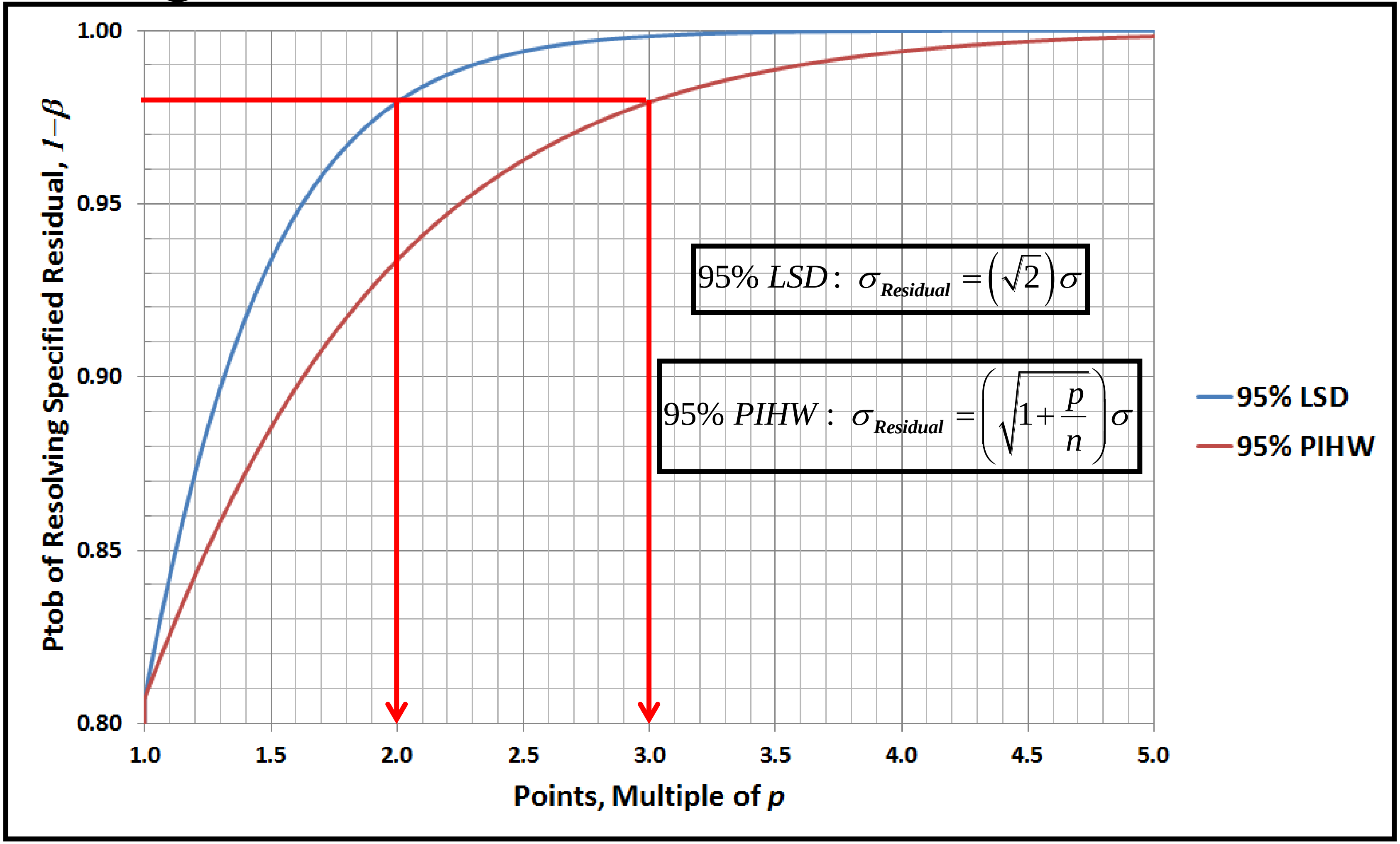
$$n = \left[\boxed{\left(\frac{z_\alpha + z_\beta}{2}\right)^2} - 1\right] p$$

Model Term-Count Multiplier

Minimum to Resolve 95% LSD or 95% PIHW with $\alpha = 0.05$



95% PIHW Tolerance Criterion is More Stringent than the 95% LSD Criterion



Numerical Scaling Example

Typical OFAT Wind Tunnel Test

- **Consider a wind tunnel test in which forces and moments are to be estimated as a function of four factors**
  - **Angles of Attack and Sideslip**
  - **Mach Number**
  - **Height (for ground effects)**
- **Typical OFAT levels might be as follows**
  - **AoA: -5^o to +15^o in 1^o increments (21 levels)**
  - **Sideslip: 0^o to +10^o in 2^o increments (6 levels)**
  - **Mach Number from 0.70 to 0.90 in 0.25 increments (9 levels)**
  - **Height (5 levels)**
- **Total of 21 x 6 x 9 x 5 = 5670 points (not atypical for OFAT test)**
- **Standard error in response estimate: σ**

Numerical Scaling Example

Corresponding MDOE Scaling Case

- **Assume adequate fits can be achieved over three AoA sub-ranges and two sideslip sub-ranges with 4th-order models**
- **A d^{th}-order model in k factors has p terms (including intercept), where**

$$p = \frac{(d+k)!}{d!k!} = \frac{(4+4)!}{4!4!} = 70$$

- **Assume a 95% LSD tolerance specification:**

$$\boxed{n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2 p}$$

Numerical Scaling Example, Cont.

- **Specify inference error risk tolerances**
  - **Max acceptable probability of rejecting a valid model: α = 0.05**
  - **Max acceptable probability of validating a bad model: β = 0.01**
- **Look up corresponding standard normal deviates, $z_{\alpha,\beta}$**
  - **For α = 0.05, z_α = 1.960 (double-sided null hypothesis)**
  - **For β = 0.01, z_β = 2.326 (single-sided alternative hypothesis)**
- **Estimate data volume per subspace:**

$$n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2 p = \frac{1}{2}\left(\frac{1.960 + 2.326}{2}\right)^2 70 = 2.296\,p = 161$$

- **Estimate total data volume (six subspaces): 6 x 161 = 966**

MDOE/OFAT Comparison

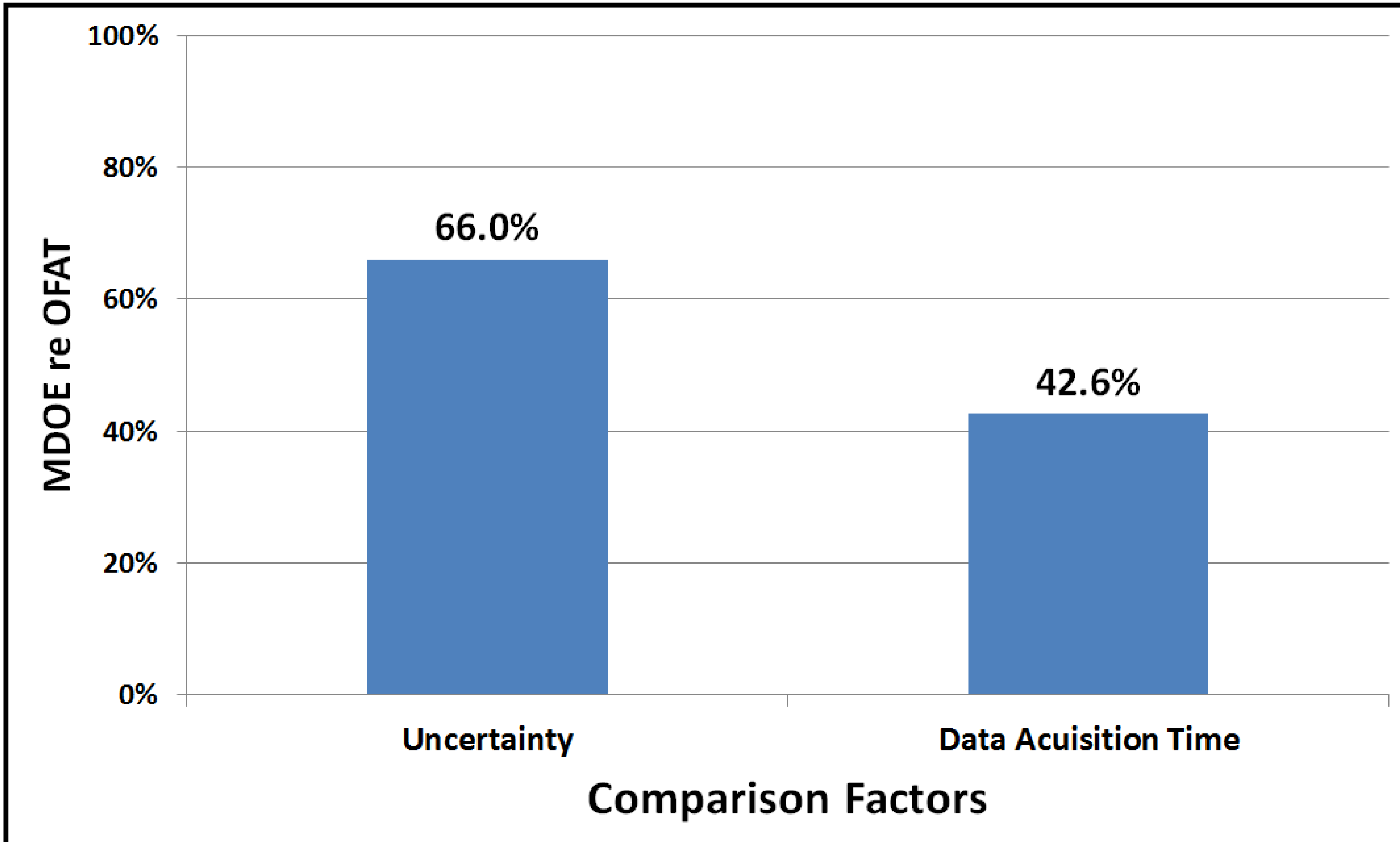
- **There is a large apparent difference in OFAT and MDOE resource requirements**
  - **OFAT: 5670 points**
  - **MDOE: 966 points**
- **The savings are not that dramatic, however**
- **MDOE methods invoke certain quality assurance tactics to defend against *covariate effects***
- **Covariates are slowly varying, persisting factors that are not controlled by the experimenter**
  - **They are generally larger than ordinary random variations**
  - **They are not reproducible from test to test**
  - **They are largely overlooked in conventional OFAT testing**

MDOE/OFAT Comparison, Cont.

- **MDOE quality assurance tactics to defend against covariates cost a factor of 1.5 to 2.5 in data rate**
  - **In the time needed to acquire 966 MDOE points, up to 2.5 times as many OFAT points might be acquired**
  - **This would be 2.5 x 966 = 2415 points**
- **The MDOE data acquisition time is thus expected to be no more than a factor of 2415/5670 of the OFAT requirement, or 42.6% (and could be rather less)**
- **Note that the scaling resulted in a data volume requirement of *n = 2.296p***
- **The MDOE standard error is thus**

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma = \left(\sqrt{\frac{p}{2.296p}}\right)\sigma = 0.660\sigma$$

Quality and Productivity Comparison



Concluding Remarks

- Quality in wind tunnel testing is more properly expressed in terms of inference error probability than unexplained variance in the raw data
  - It is more important to get the *right answer*, than "good data"
  - This imposes a responsibility to articulate tolerance requirements
- There is a mathematical relationship between resource requirements and quality requirements
- Each new data point reduces inference error risk
  - Too little data means unacceptable inference error risk
  - Too much data means wasted resources
- The experimental aeronautics community might consider adopting the 95% LSD as a tolerance specification
- Then data volume in the range of 2 to 3 times the number of points needed to fit a model would typically be sufficient

NASA

NAVAL
POSTGRADUATE
SCHOOL

Hybrid Designs: Space Filling and Optimal Experimental Designs for Use in Studying Computer Simulation Models

Rachel Johnson Silvestrini PhD
Assistant Professor
Operations Research Department

Outline

- Computer Simulations
- Metamodeling
- Types of Experimental Designs
- A Hybrid Experimental Design Approach
- Example Using a ISR Simulation Model
- Questions

Computer Simulations

- Computer simulation models are built to mimic reality
- We ***do not*** always treat computer simulations like reality
  - Lots and lots of experiments are run
  - Many experiments are run that are prohibited in the real world
- We ***do*** often treat computer simulation results as if they were reality

Types of Simulation Models

- Stochastic
  - Output is a random variable
  - Blocking and randomization not an issue, but replication is
- Deterministic
  - For a given set of inputs, the output will be the same each time the model is run
  - Blocking, randomization, and replication are irrelevant

Illustration and Definitions

Noise, unknown factors, model misspecification

Inputs: x_i (variables that you can control)

Your Simulation or Model

Output(s): y (variable that you want to measure)

- Let's assume that there is some underlying model: $y = f(x) + \varepsilon$
- This model (often called a metamodel) can be:
    - mechanistic
    - empirical
        - linear regression model
        - non-linear model
        - generalized linear model
        - Gaussian process model (aka Kriging)
- A goal of *Experimental Design*: find/fit a metamodel

Simulation Needs

- People who run simulation models sometimes have trouble choosing what conditions to run (input levels to select) in order to fully characterize the input domain
- Additionally, once those conditions are selected, it might be difficult to describe the outputs in a meaningful way
- Experimental design and analysis provides a way to
  - Choose conditions to run your model (i.e. select inputs)
  - Find a suitable mathematical model that allows you to summarize your input-output data

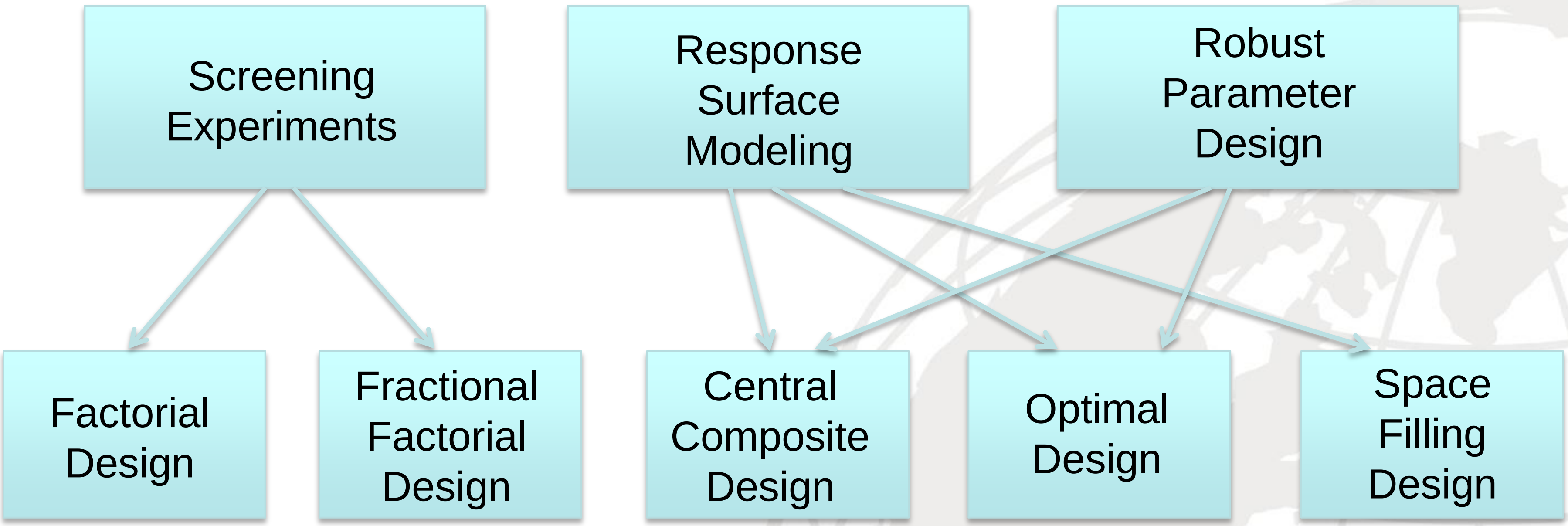
Metamodeling: What is it?

- After we run a computer simulation we would like to relate the inputs to the output(s) through the use of a closed form mathematical expression [1,4]
- Examples of common empirical metamodels used in practice:
  - Linear regression models (i.e. polynomial models) [1,4]
  - Non-linear models (i.e. logistic regression model) [8]
  - Gaussian Process models (i.e. Kriging) [4,9]

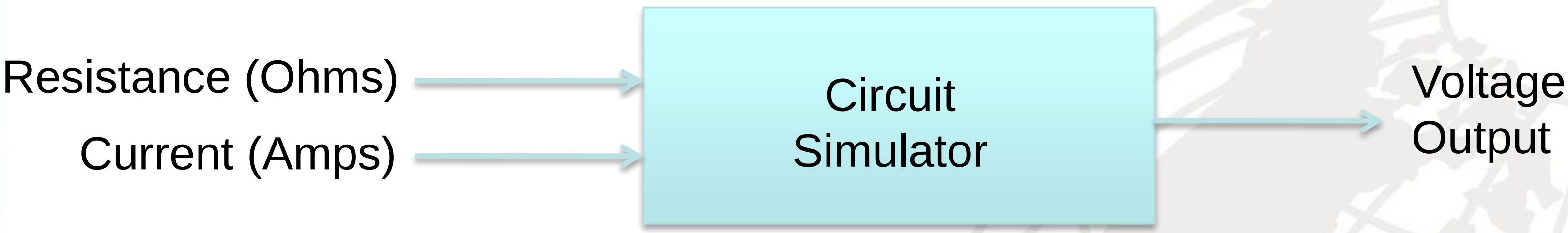
Experimental Designs

- The choice of experimental design can strongly influence how “good” your results are
- Can create most designs in standard software packages such as JMP or Design Expert
- Require you to list
  - Inputs (including input values or ranges)
  - Output(s)
  - Number of runs (trials you are willing to perform)
  - If running optimal design – must specify assumed model

Experimental Designs are Based on Goals

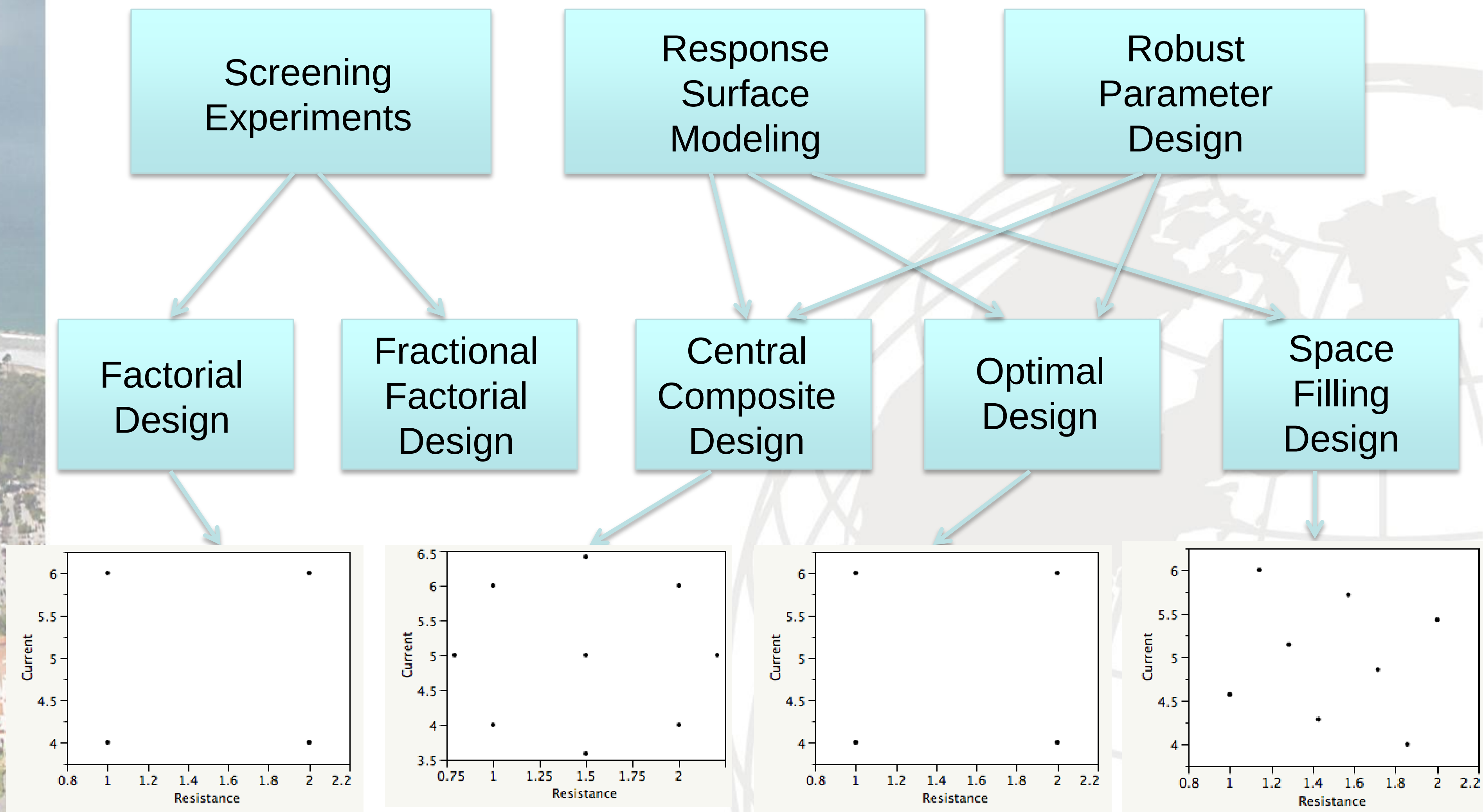


A Quick Example: Studying Ohm's Law



Input Factor	Range	Variable Type
Resistance (ohms)	[1 – 2]	Continuous
Current (amps)	[4 - 6]	Continuous

Experimental Designs are Based on Goals



Optimal Experimental Design

Optimal Design (examples: *D*-optimal and *I*-optimal)

- Pros
  - Great for creating empirical models of many forms (especially useful if using the linear regression approach)
  - Useful for constrained design spaces
  - Optimal designs for many linear regression models are the standard designs (i.e. 2^k)
- Cons
  - Requires specification of the metamodel before collecting any data
  - Non-linear optimal designs are dependent on unknown parameters

Space Filling Experimental Design

Space Filling Design (examples: Latin Hypercube and Uniform)

- Pros
  - Fill the design space
  - Useful for unknown metamodel choices
- Cons
  - Don't cover the corners of the design space

Experimental Design Approach

- What if you don't know *a priori* what type of metamodel will work best for your results?
- Is there some type of experimental design that can be used assuming you may choose several type of metamodels to use?
  – Yes
  – A hybrid design approach that combines optimal design with space filling design
    - Provides coverage to the corners of the design space and the interior
    - Useful for fitting linear regression models and for fitting metamodels you might not have planned for

Situations Useful for Hybrid Design

- Situation 1: You are running a simulation experiment and would like good coverage of the design space. You are not sure what metamodel you will use, but think that a linear regression model choice is among the possibilities

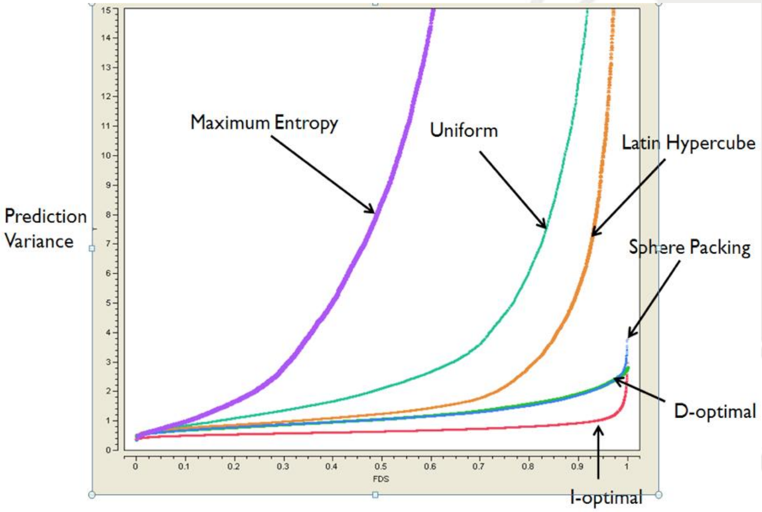
- Situation 2: You are running a simulation experiment and will most likely fit a linear regression model, but would like to simulation some "random" trials to use as either
  - Cross-validation or
  - If your model is making bad predictions, points that can be used to fit new models

Research Comparison Method

- Previous research on these designs compared optimal, space filling designs, and hybrid designs based on:
  - Scaled prediction variance
    - For the linear regression model: $\frac{N\,V[\hat{y}(\boldsymbol{x}_o)]}{\sigma^2} = N\boldsymbol{x}_o'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_o$
    - For the Gaussian process mode:
  - Fraction of Design Space plots [11]: plot the empirical distribution function of scaled prediction variance over the design region
  - Used in the assessment of prediction capability [8]

Example FDS Plot

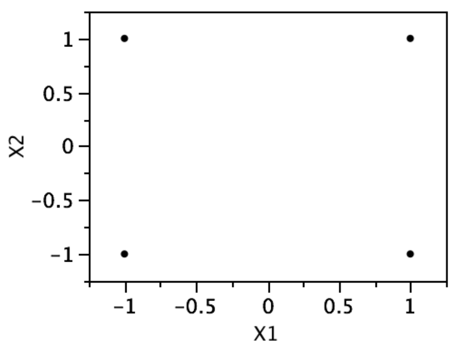
- Designs created for the case with
  - 2 input factors
  - 2^{nd} order polynomial
  - Sample size = 10



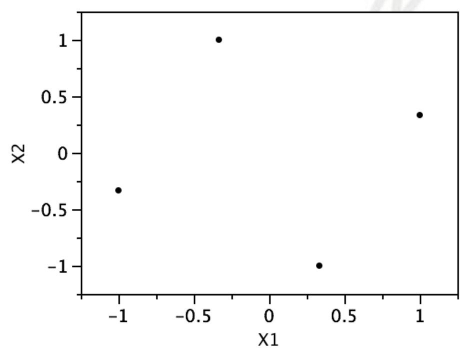
Hybrid Design Development

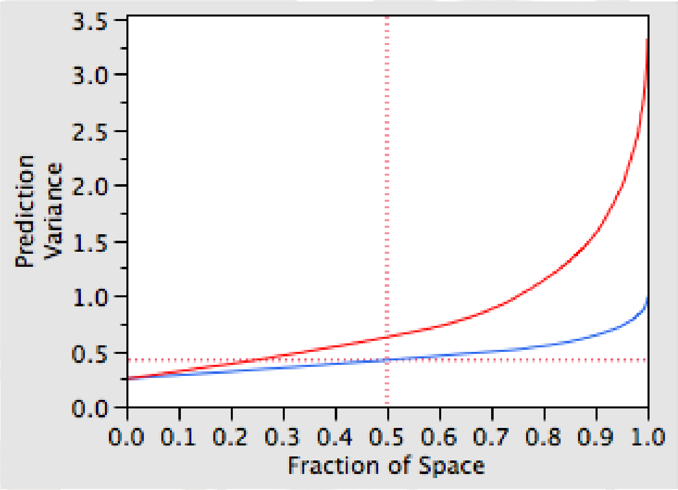
- Consider a saturated design for two factors and an anticipated main effects and two factor interaction model
- Here is an example of what the I-optimal design (a) and a Latin hypercube space-filling design (b) look like and their associated FDS plots

a) *I*-optimal



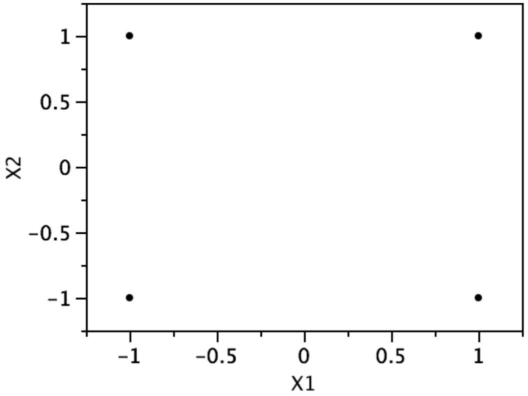
b) Latin hypercube

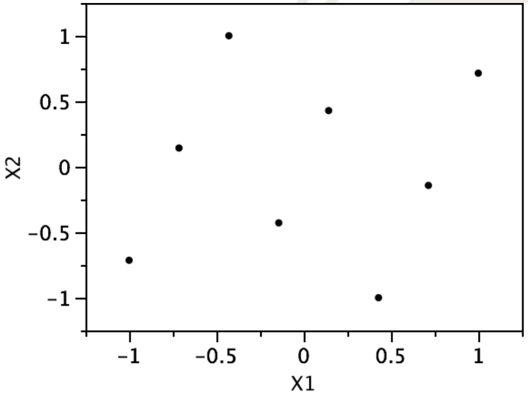


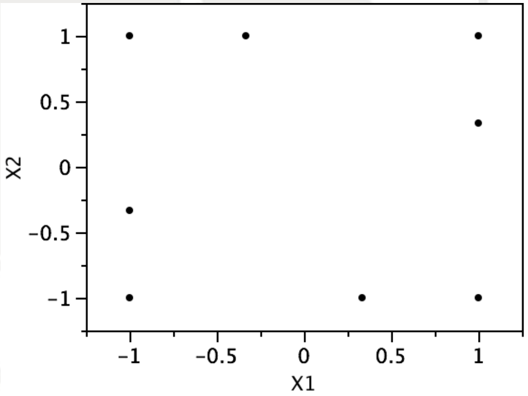


Hybrid Design Points

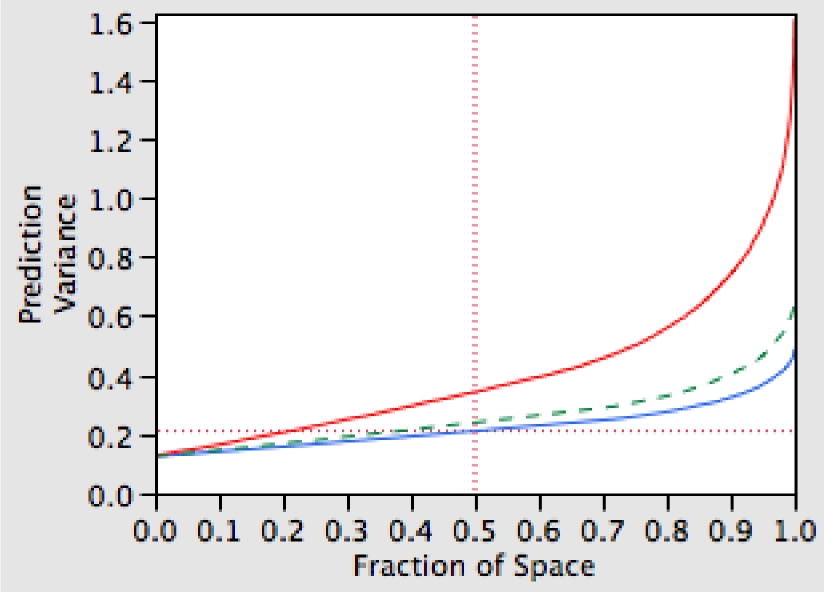
- We augmented the space-filling design with optimal points
- Why?
  - Wanted the space-filling design because it fills the interior region of the design space
  - Wanted the optimal design points







Performance of Hybrid Design



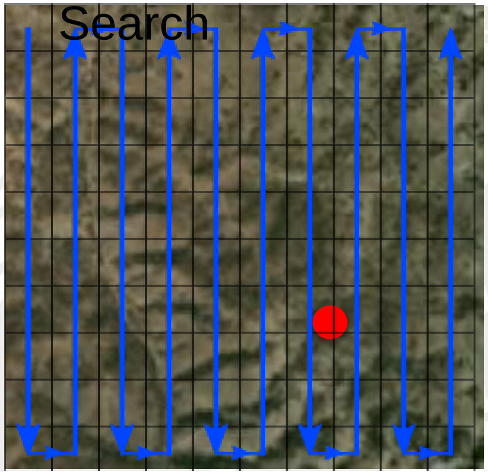
ISR Application

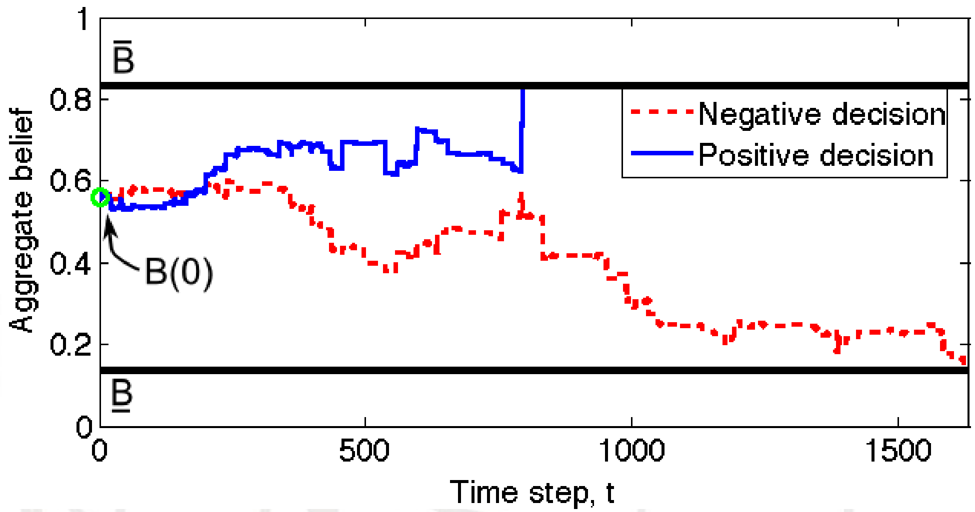
- Unmanned systems can play a prominent role in diverse information gathering missions such as:
  - Search and Rescue (SAR)
  - Intelligence, Surveillance and Reconnaissance (ISR)
- Current research on unmanned system search requires the use of sophisticated sensing, computation, coordination, and communication capabilities
- This example is based on research conducted by a colleague (Timothy H. Chung) and I
- The work presented seeks to revisit the used of exhaustive search strategies as the basis of the search process and leverage new probability models as well as experimental design to help inform and refine concepts of operations

The ISR Simulation

- Consider an area of interest with a missing person or a target
- A simulator was built to mimic an unmanned system searching the area
- The simulator updates probabilities about the location of the person as a function of time and observation
- The goal of the study is to study the effect of several inputs of interest on five response variables

Lawnmower Search



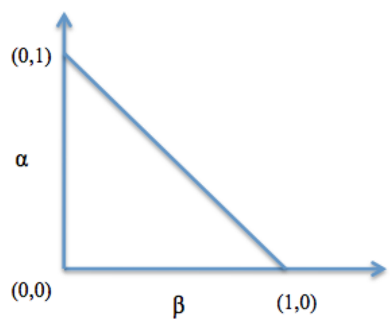


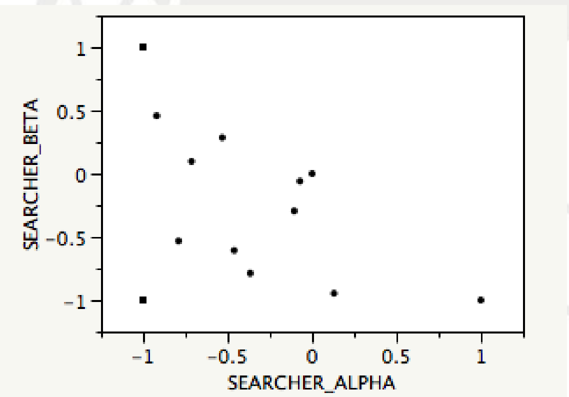
Hybrid Design Approach

- Combined D-optimal design points with uniform design points
- The inputs are:

Factor	Label	Description	Range or Levels
α	x_1	False positive detection error	[0.0, 1.0]
β	x_2	False negative detection error	[0.0, 1.0]
$B(0)$	x_3	Initial aggregate belief	[0.3, 0.7]
$\overline{B}$	x_4	Upper decision threshold	[0.8, 0.95]
$\underline{B}$	x_5	Lower decision threshold	[0.05, 0.2]
$\mathcal{M}(0)$	x_6	Initial target probability map	{good, bad, none}
SP	x_7	Exhaustive search pattern	{lawnmower, sweeping}

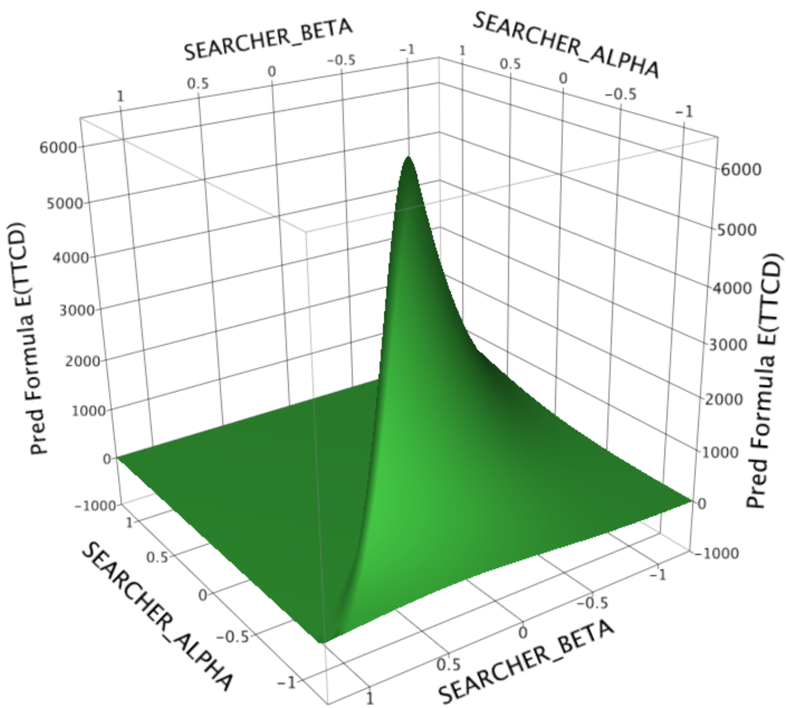
- A picture of the hybrid D-optimal and uniform design in two of the factors is illustrated as

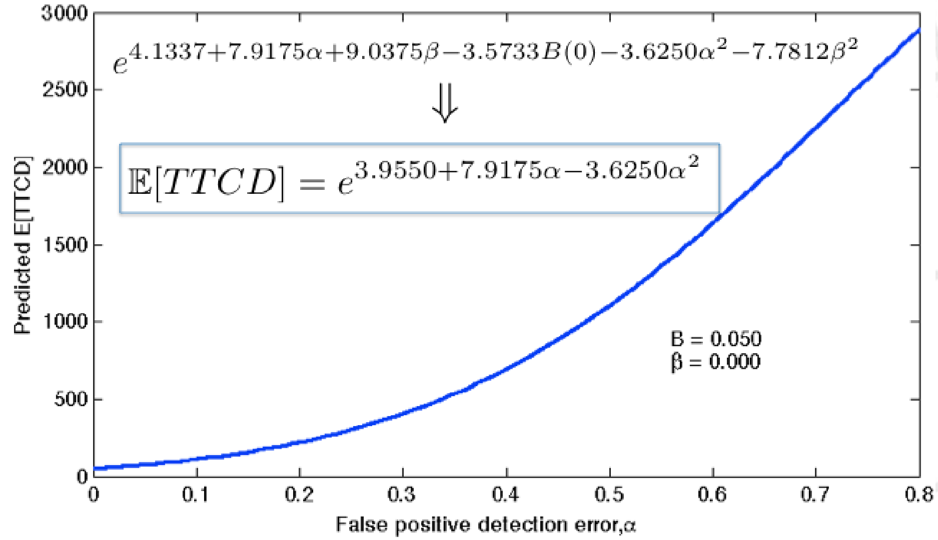




Results

Lawnmower	Factors														
		Main Effects							Interactions					Squared Effects	
		x_1	x_2	x_3	x_5	x_6			x_1x_3	x_1x_6			x_3x_5	x_1^2	x_2^2
Response	Intercept					*bad*	*good*	*none*		*bad*	*good*	*none*			
% Correct Neg	-0.147	-0.172		-0.888											
% Correct Pos	-0.917	-0.094		1.163	-0.404	-0.657	0.330	0.330		-0.436	0.198	0.238	0.271		
E[TTCD]*	8.644	2.007	1.125		-0.268									-0.580	-1.245
E[TTCND]*	8.462	1.800	1.175	0.415	-0.319										-1.541
E[TTCPD]*	8.737	2.262	0.961	0.110					-0.278					-0.533	-1.549





References

- [1] Allen, T.T., Bernshteyn, M.A., and Kabiri-Bamoradian, K. (2003). "Constructing Meta-Models for Computer Experiments," *Journal of Quality Technology* **35**(3), pp. 264 – 274.
- [2] Atkinson, A. and Donev, A. (1992). *Optimum Experimental Designs*, Oxford Science Publications, Oxford.
- [3] Fang, K.T. (1980). "The Uniform Design: Application of Number-Theoretic Methods in Experimental Design," *Acta Math. Appl. Sinica.* **3**, pp. 363 – 372.
- [4] Hussain, M.F., Barton, R.R., and Joshi, S.B. (2002). "Metamodeling: Radial Basis Functions, Versus Polynomials," *European Journal of Operational Research* **138**, pp. 142 – 154.
- [5] Johnson, M.E., Moore, L.M. and Ylvisaker, D. (1990). "Minimax and maxmin distance design," *Journal for Statistical Planning and Inference* **26**, pp. 131 – 148.
- [6] Johnson, R.T., Montgomery, D.C., Jones, B., and Parker, P. (2010). "Comparing Computer Experiments for Fitting High Order Polynomial Metamodels," *Journal of Quality Technology*,42(1), pp. 86 – 102.
- [7] McKay, N. D., Conover, W. J., Beckman, R. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics* **21**, pp. 239 – 245.
- [8] Myers, R.H. and Montgomery, D.C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd ed. John Wiley, New York, NY.
- [9] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments," *Statistical Science* **4**(4), pp. 409 – 423.
- [10] Shewry, M.C. and Wynn, H.P. (1987). "Maximum entropy sampling," *Journal of Applied Statistics* **14**, pp. 898 – 914.
- [11] Zahran, A., Anderson-Cook, C.M., and Myers, R.H. (2003). "Fraction of Design Space to Assess the Prediction Capability of Response Surface Designs," *Journal of Quality Technology* **35**, pp. 377 – 386.

Beyond Probability

A pragmatic approach to uncertainty quantification in engineering

Scott Ferson, Applied Biomathematics, scott@ramas.com

NASA Statistical Engineering Symposium, Williamsburg, Virginia, 4 May 2011

Wishful thinking

- Using inputs or models because they are convenient, or because you hope they're true

Kansai International Airport

- 30 km from Kobe in Osaka Bay
- Artificial island made with fill
- Engineers told planners it'd sink [6, 8] m
- Planners elected to design for 6 m
- It's sunk 9 m so far and is still sinking

(The operator of the airport denies these media reports)

Variability = aleatory uncertainty

- Arises from natural stochasticity

- Variability arises from
  - spatial variation
  - temporal fluctuations
  - manufacturing or genetic differences

- Not reducible by empirical effort

Incertitude = epistemic uncertainty

- Arises from incomplete knowledge

- Incertitude arises from
  - limited sample size
  - mensurational limits (‘measurement error’)
  - use of surrogate data

- Reducible with empirical effort

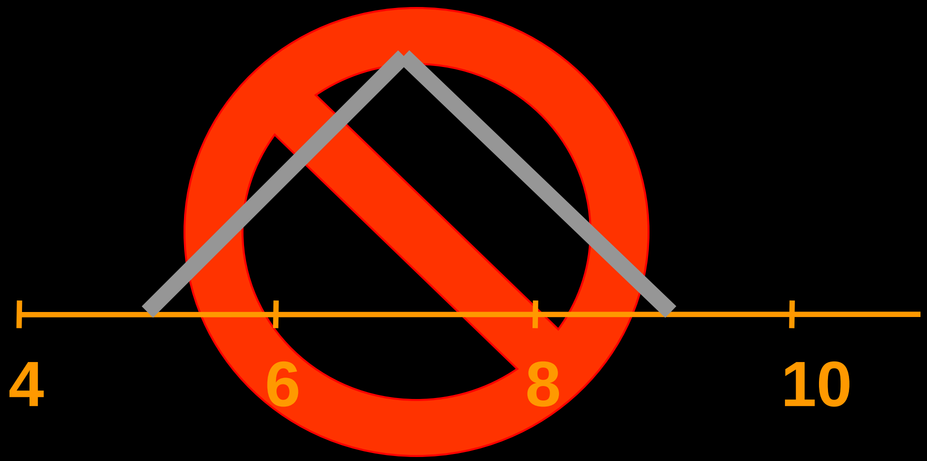
Propagating incertitude

Suppose

A is in [2, 4]

B is in [3, 5]

What can be said about the sum $A+B$?



The right answer for engineering is [5,9]

They must be treated *differently*

- Variability should be modeled as randomness with the methods of probability theory
- Incertitude should be modeled as ignorance with the methods of interval analysis
- Imprecise probabilities can do both at once

Incertitude is common in engineering

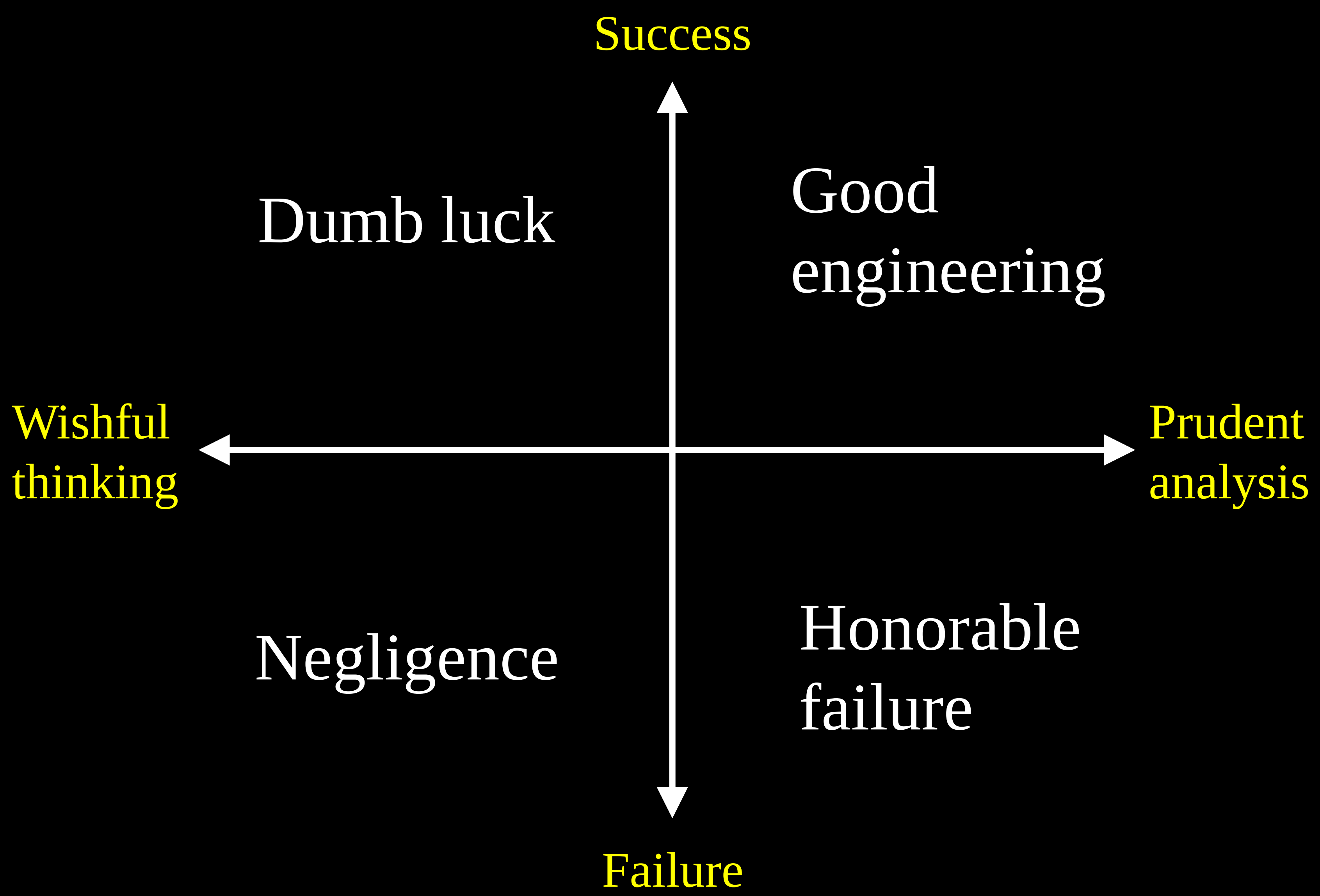
- Periodic observations
  When did the fish in my aquarium die during the night?
- Plus-or-minus measurement uncertainties
  Coarse measurements, measurements from digital readouts
- Non-detects and data censoring
  Chemical detection limits, studies prematurely terminated
- Privacy requirements
  Epidemiological or medical information, census data
- Theoretical constraints
  Concentrations, solubilities, probabilities, survival rates
- Bounding studies
  Presumed or hypothetical limits in what-if calculations

Wishful thinking

- Pretending you know the
  - Value
  - Distribution function
  - Dependence
  - Model

  when you don't is wishful thinking

- Uncertainty analysis makes a prudent analysis


Success
Dumb luck
Good engineering
Wishful thinking
Prudent analysis
Negligence
Honorable failure
Failure

Traditional uncertainty analyses

- Interval analysis
- Taylor series approximations (delta method)
- Normal theory propagation (ISO/NIST)
- Monte Carlo simulation
- Stochastic PDEs
- Two-dimensional Monte Carlo

Untenable assumptions

- Uncertainties are small
- Distribution shapes are known
- Sources of variation are independent
- Uncertainties cancel each other out
- Linearized models good enough
- Underlying physics is known and modeled

Need ways to relax assumptions

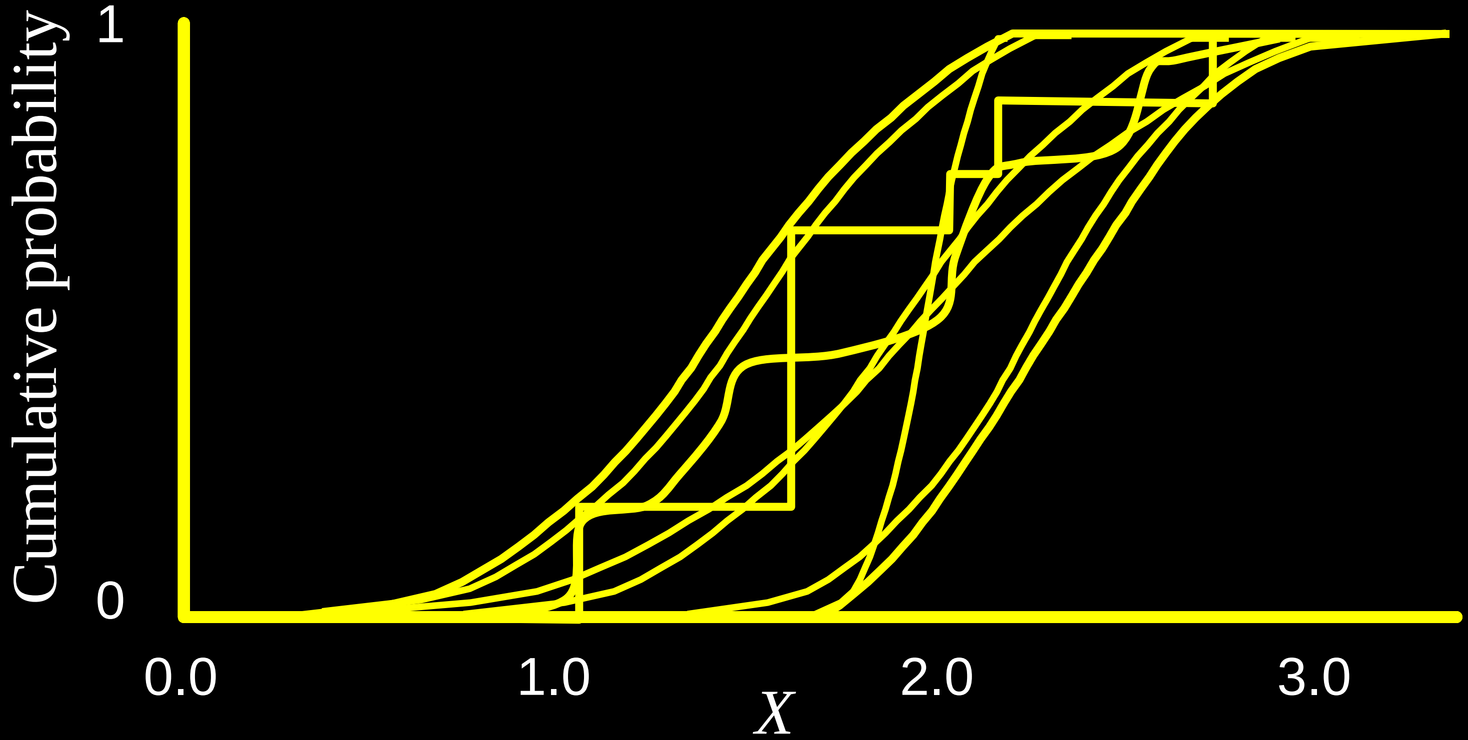
- Hard to say what the distribution is precisely
- Non-independent, or *unknown* dependencies
- Uncertainties that may not cancel
- Possibly large uncertainties
- Model uncertainty

Probability bounds analysis (PBA)

- Sidesteps the major criticisms
  - Doesn’t force you to make any assumptions
  - Can use only whatever information is available
- Bridges worst case and probabilistic analysis
- Distinguishes variability and incertitude
- Acceptable to both Bayesians and frequentists

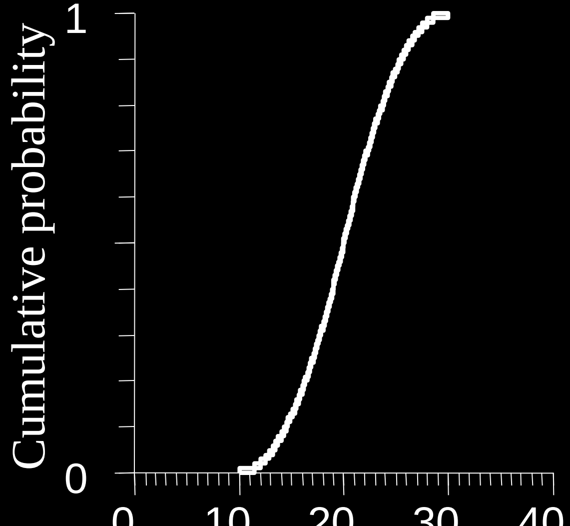
Probability box (p-box)

Interval bounds on a cumulative distribution function

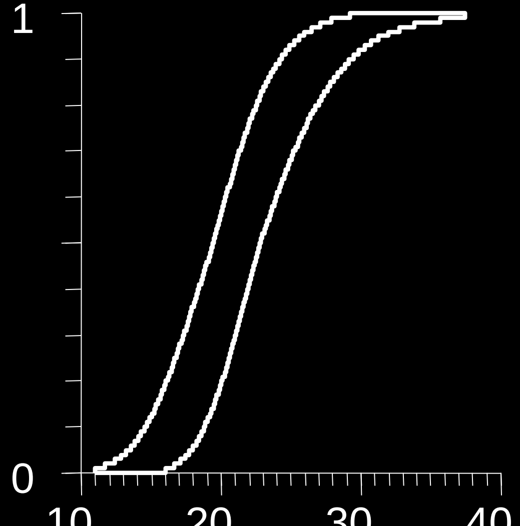


Uncertain numbers

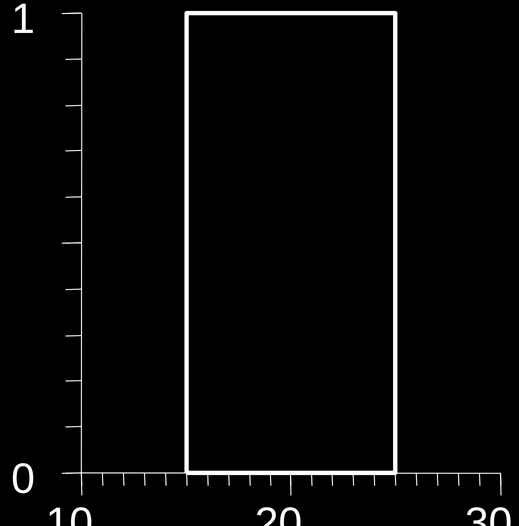
Probability distribution



Probability box



Interval



Not a uniform distribution

Uncertainty arithmetic

- We can do math on p-boxes
- When inputs are distributions, the answers conform with probability theory
- When inputs are intervals, the results agree with interval (worst case) analysis

Calculations

- All standard mathematical operations
  - Arithmetic (+, –,   ,   , ^, min, max)
  - Transformations (exp, ln, sin, tan, abs, sqrt, etc.)
  - Magnitude comparisons (<, ≤, >, ≥, ⊆)
  - Other operations (nonlinear ODEs, finite-element methods)
- Faster than Monte Carlo
- Guaranteed to bound the answer
- Optimal solutions often easy to compute

Probability bounds analysis

- Special case of imprecise probabilities
- Addresses many problems in risk analysis
  - Input distributions unknown
  - Imperfectly known correlation and dependency
  - Large measurement error, censoring
  - Small sample sizes
  - Model uncertainty

Better than sensitivity analysis

- Unknown distribution is hard for sensitivity analysis since infinite-dimensional problem
- Analysts usually fall back on a maximum entropy approach, which erases uncertainty rather than propagates it
- Bounding seems very reasonable, so long as it reflects all available information

Example: uncontrolled fire

$$F = A \,\&\, B \,\&\, C \,\&\, D$$

Probability of ignition source
Probability of abundant fuel presence
Probability fire detection not timely
Probability of suppression system failure

Imperfect information

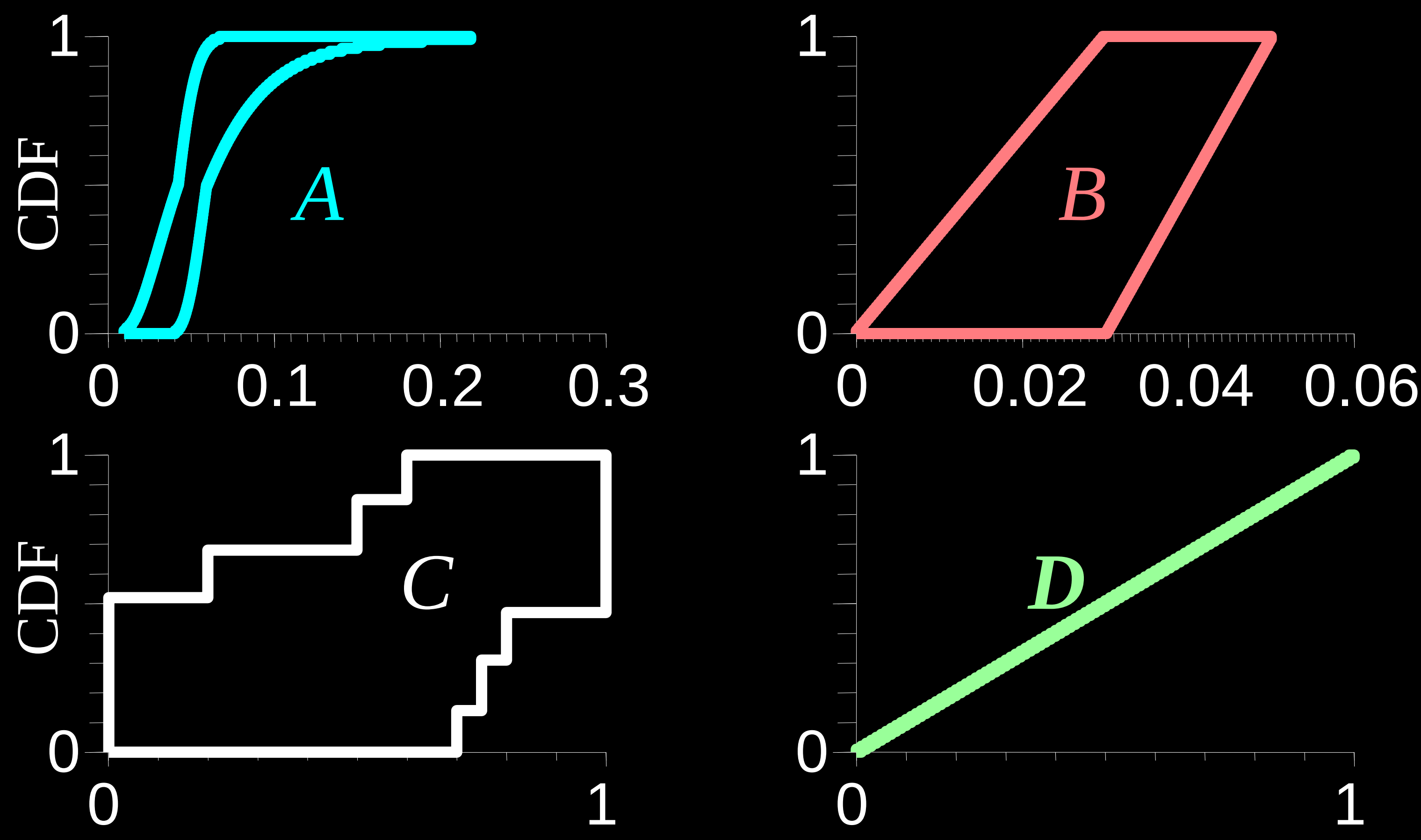
- Calculate $A \& B \& C \& D$, with partial information:
  - A's distribution is known, but not its parameters
  - B's parameters known, but not its shape
  - C has a small empirical data set
  - D is known to be a precise distribution
- Bounds assuming independence?
- Without any assumption about dependence?

A = {lognormal, mean = [.05,.06], variance = [.0001,.001])

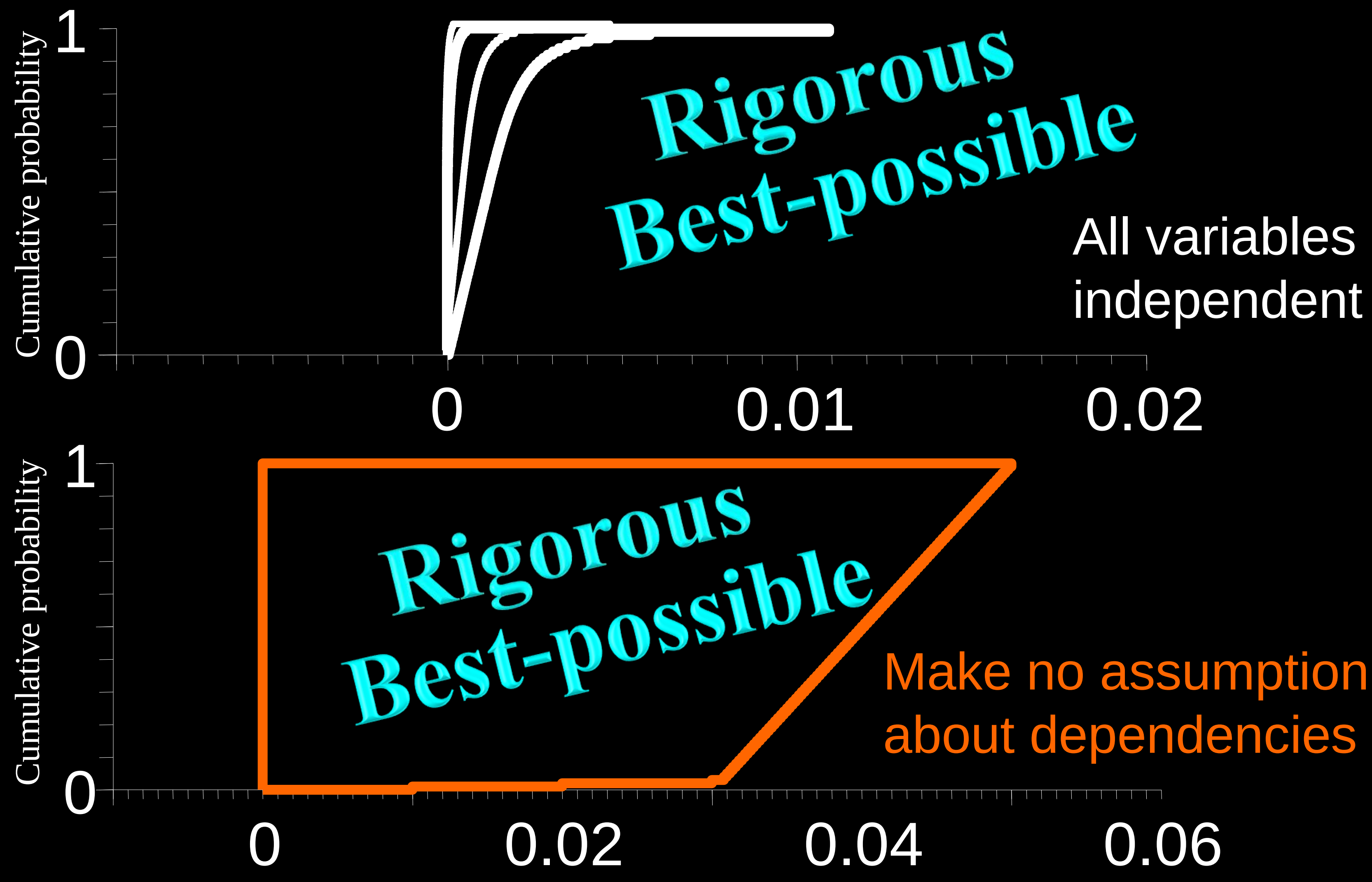
B = {min = 0, max = 0.05, mode = 0.03}

C = {sample data = 0.2, 0.5, 0.6, 0.7, 0.75, 0.8}

D = uniform(0, 1)



Resulting answers



Summary statistics

Independent	
Range	[0, 0.011]
Median	[0, 0.00113]
Mean	[0.00006, 0.00119]
Variance	[2.9×10^{-9}, 2.1×10^{-6}]
Standard deviation	[0.000054, 0.0014]
No assumptions about dependence	
Range	[0, 0.05]
Median	[0, 0.04]
Mean	[0, 0.04]
Variance	[0, 0.00052]
Standard deviation	[0, 0.023]

How to use the results

When uncertainty makes no difference
(because results are so clear), bounding gives confidence in the reliability of the decision

When uncertainty swamps the decision
(*i*) use other criteria within probability bounds, or
(*ii*) use results to identify inputs to study better

Justifying further empirical effort

- If incertitude is too wide for decisions, and bounds are best possible, more data is needed

- Strong argument for collecting more data

Advantages

- Computationally efficient
  - No simulation or parallel calculations needed
- Fewer assumptions
  - Not just different assumptions, *fewer* of them
  - Distribution-free probabilistic risk analysis
- Rigorous results
  - Built-in quality assurance
  - Automatically verified calculation

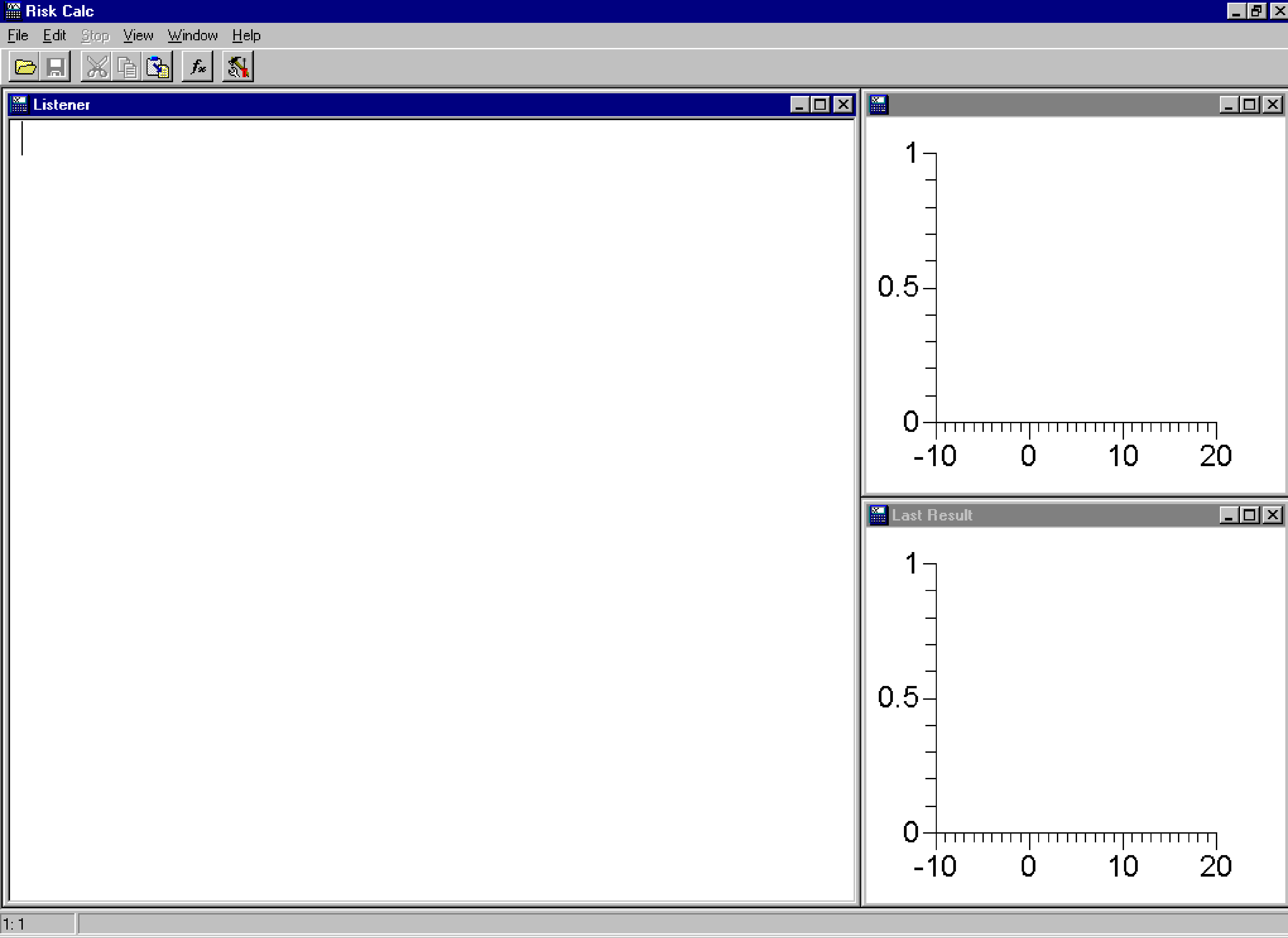
Disadvantages

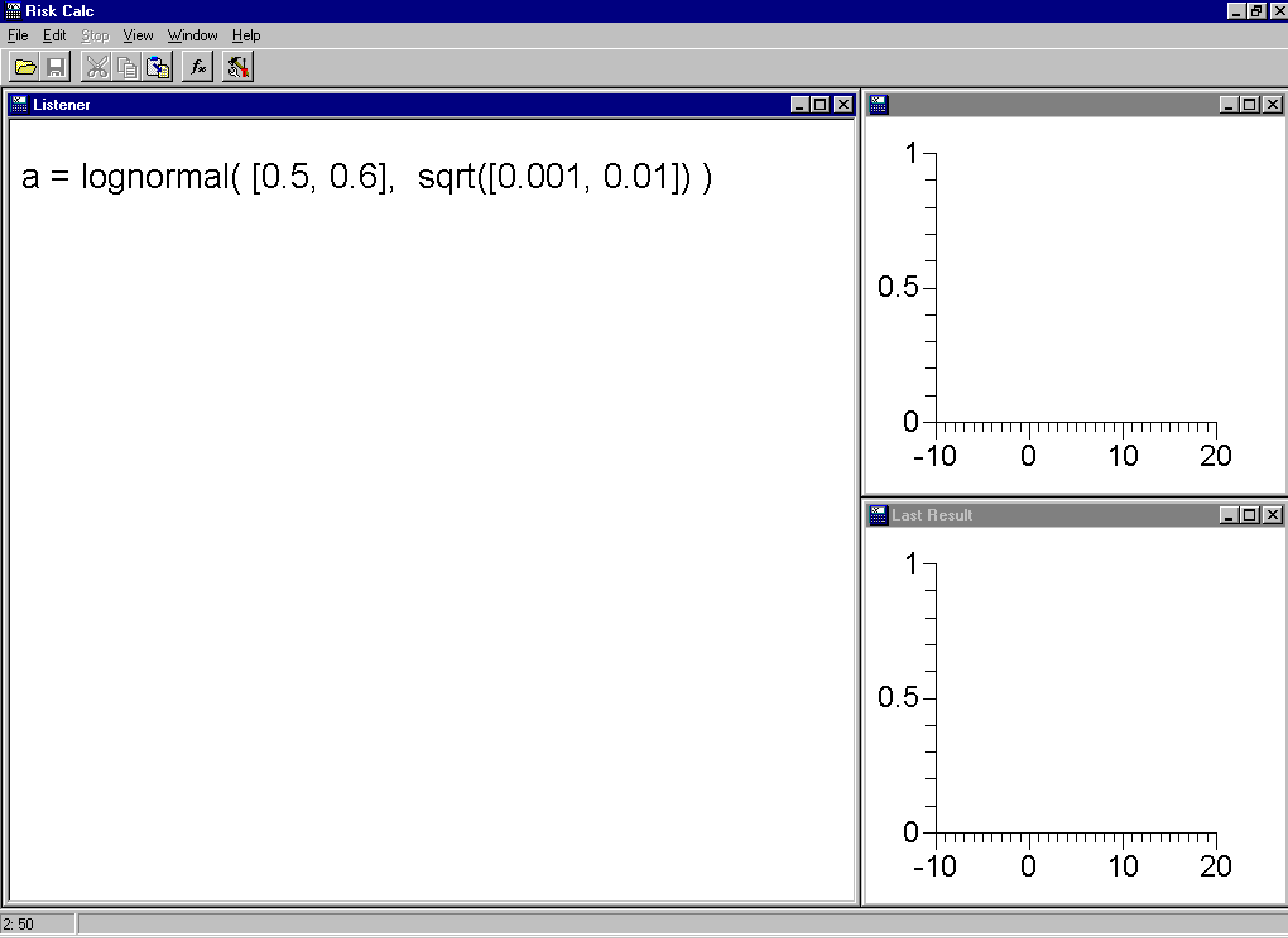
- P-boxes don't say what outcome is most likely
- Hard to get optimal bounds on non-tail risks
- Some technical limits (e.g., sensitive to repeated variables, tricky with black boxes)
- A p-box may not express the tightest possible bounds given all available information (although it often will)

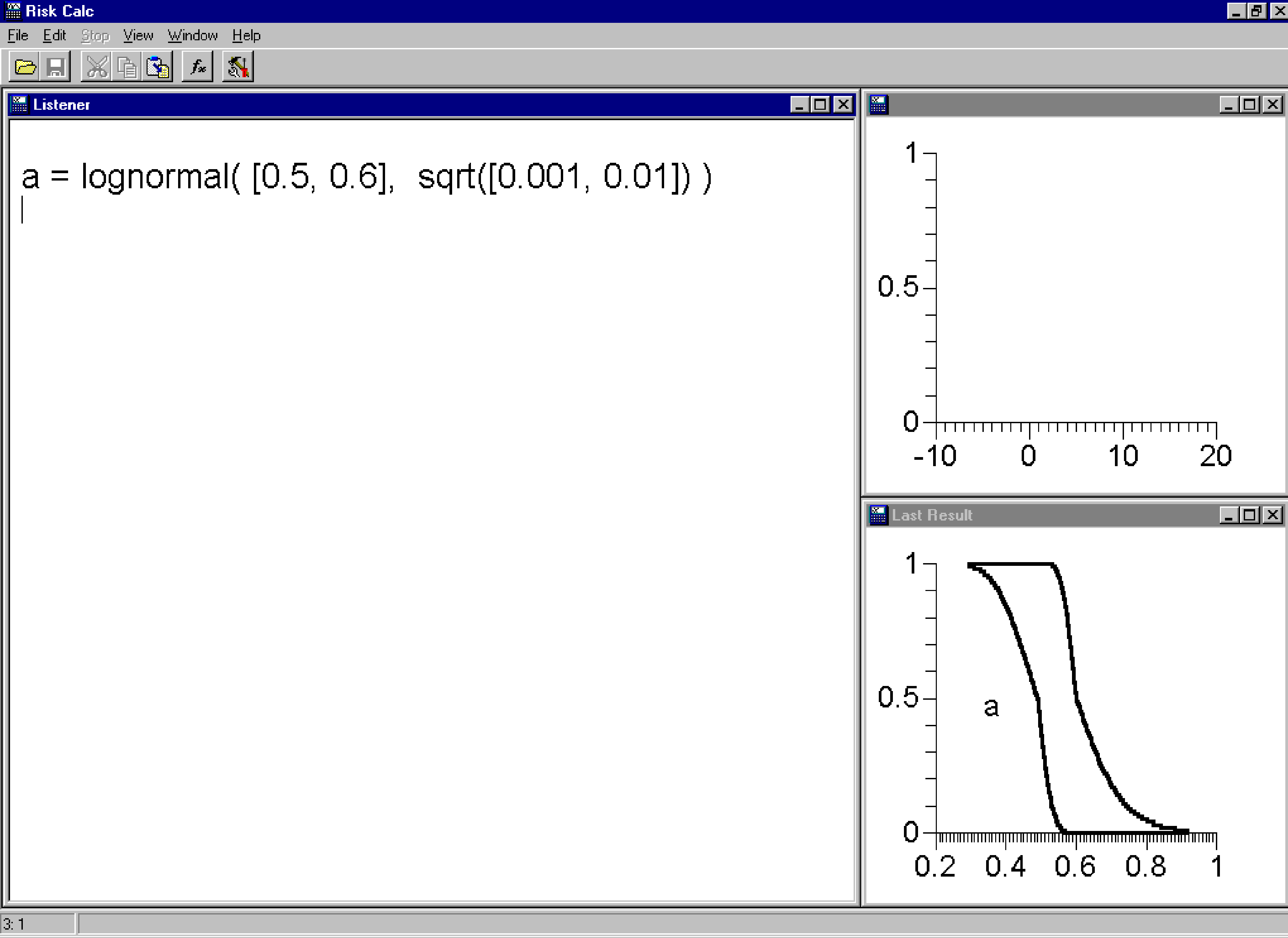
Software

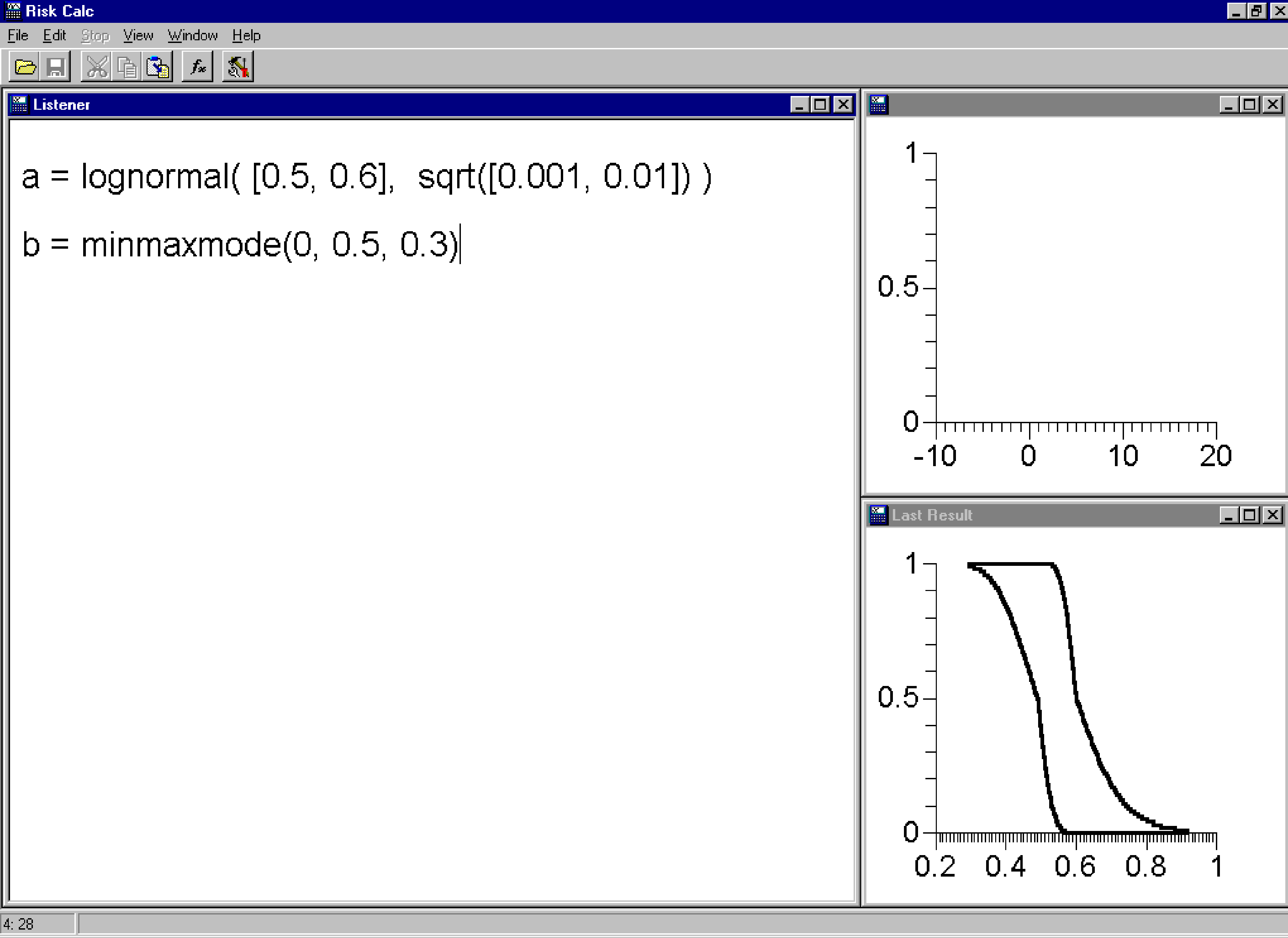
Need β testers

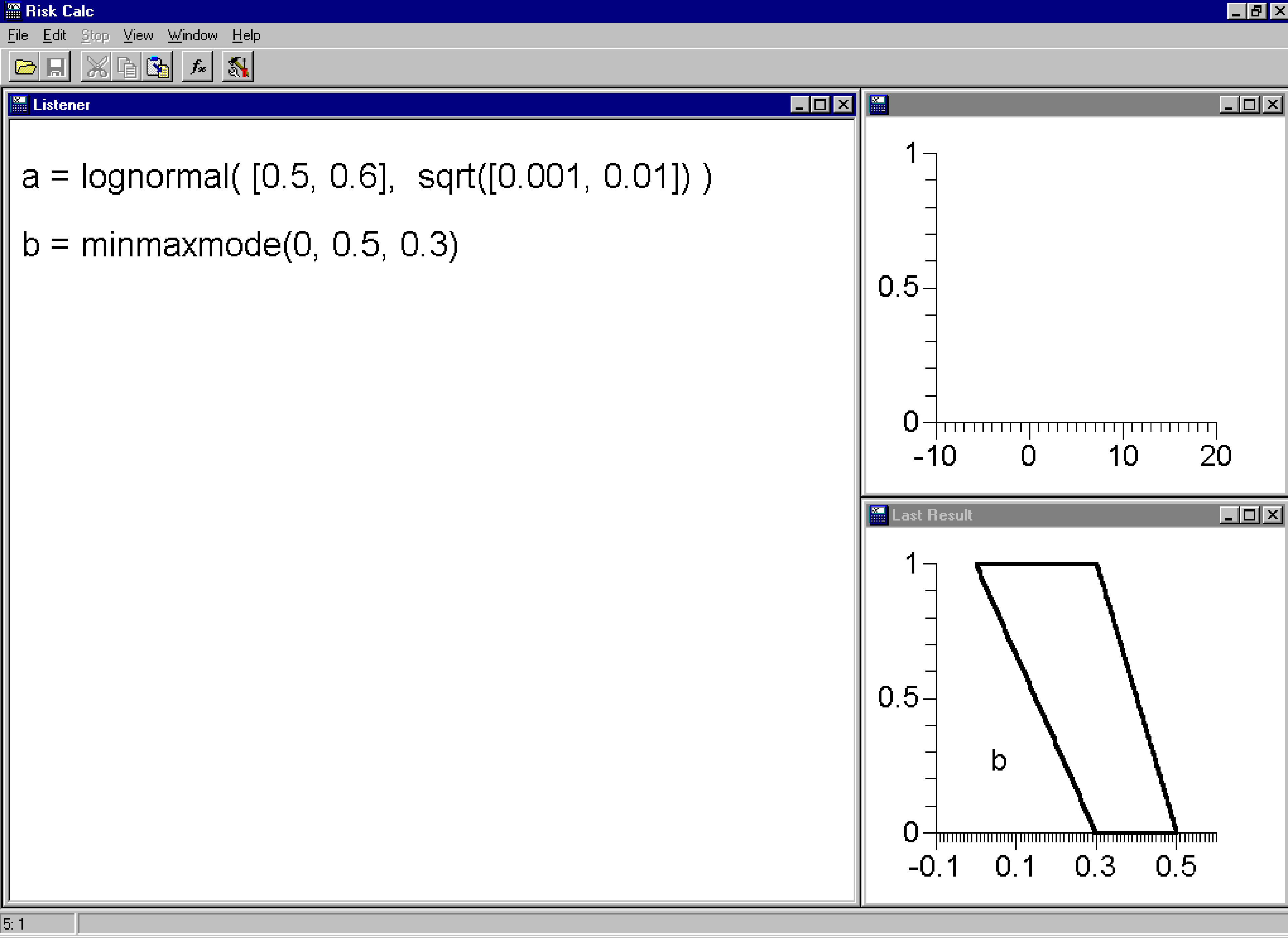
- UC add-in for Excel (NASA, beta 2011)
- RAMAS Risk Calc 4.0 (NIH, commercial)
- Statool (Dan Berleant, freeware)
- Constructor (Sandia and NIH, freeware)
- Pbox.r library for R
- PBDemo (freeware)
- Williamson and Downs (1990)

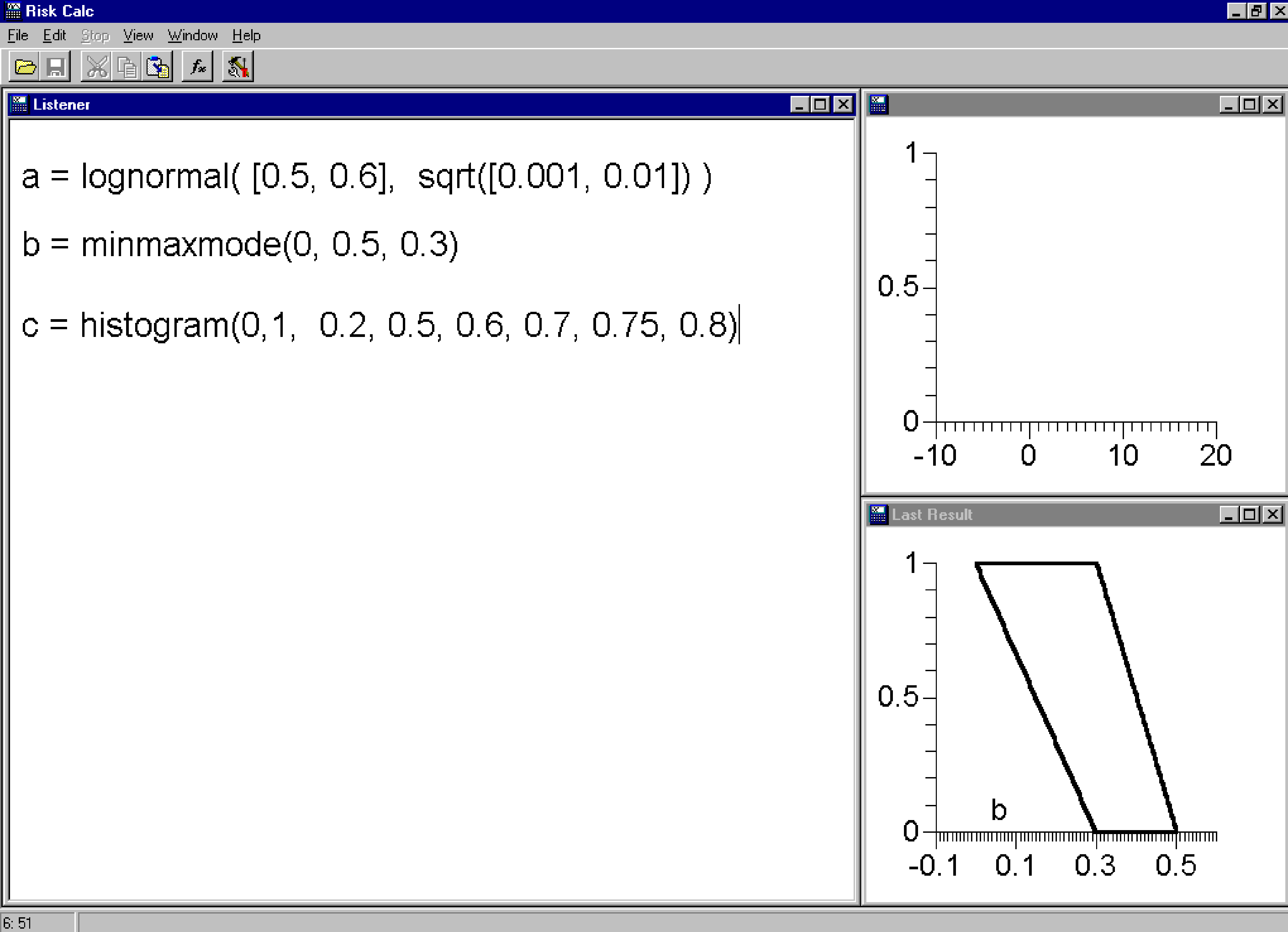

Risk Calc
File
Edit
Stop
View
Window
Help
Listener
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
-10
0
10
20
1: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
-10 0 10 20
2: 50

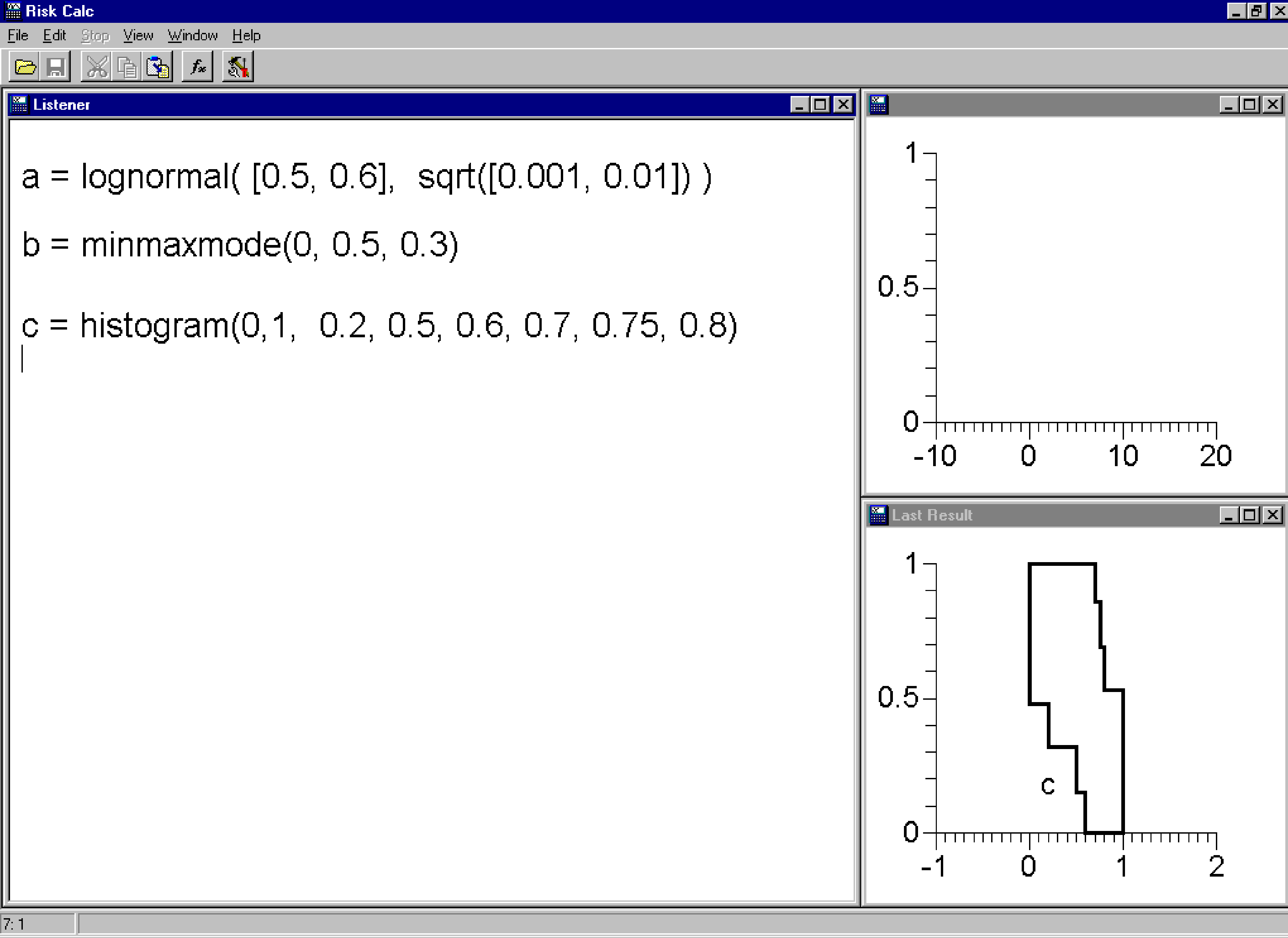

Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
a
0.2 0.4 0.6 0.8 1
3: 1

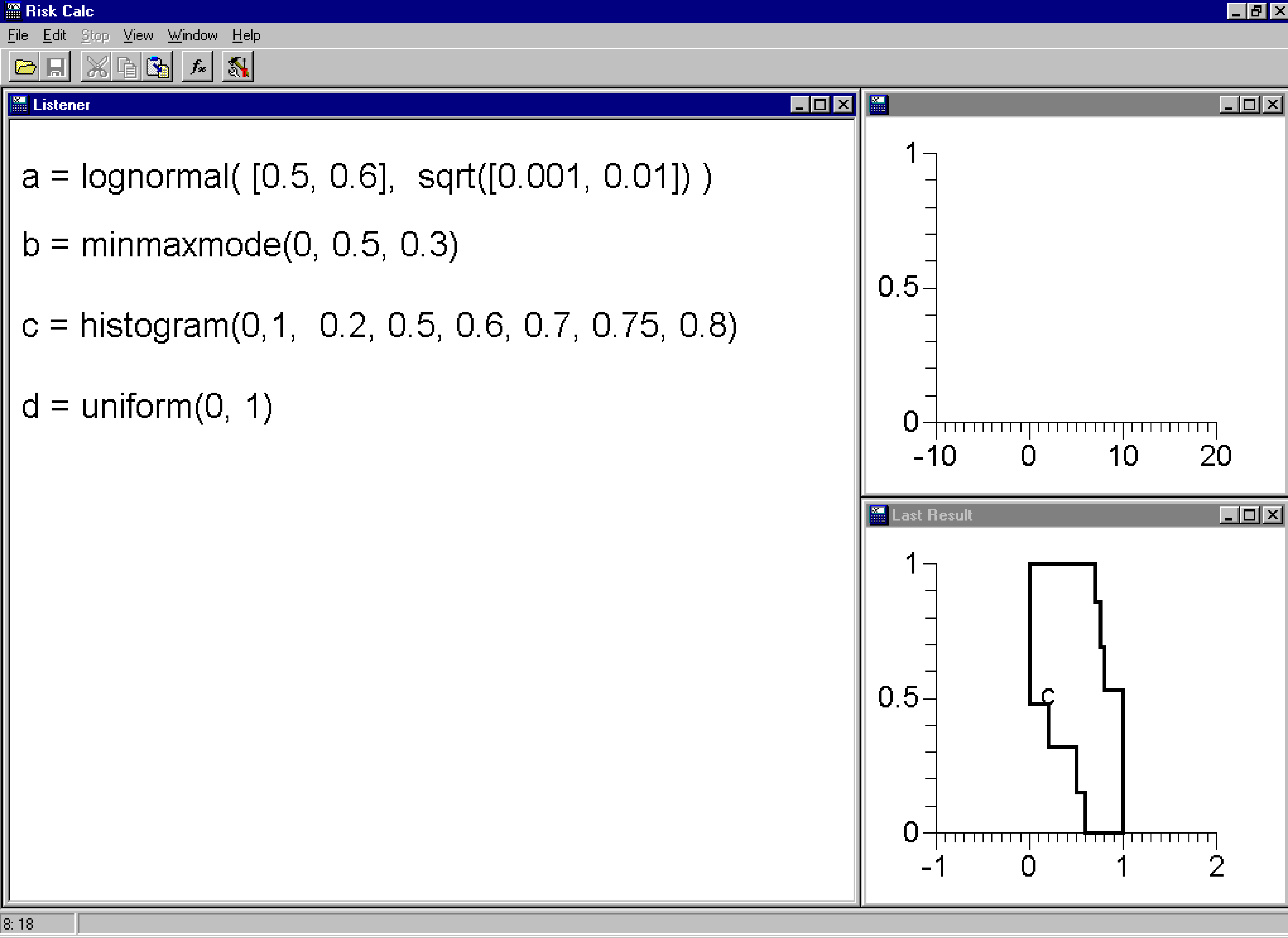

Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
0.2 0.4 0.6 0.8 1
4: 28

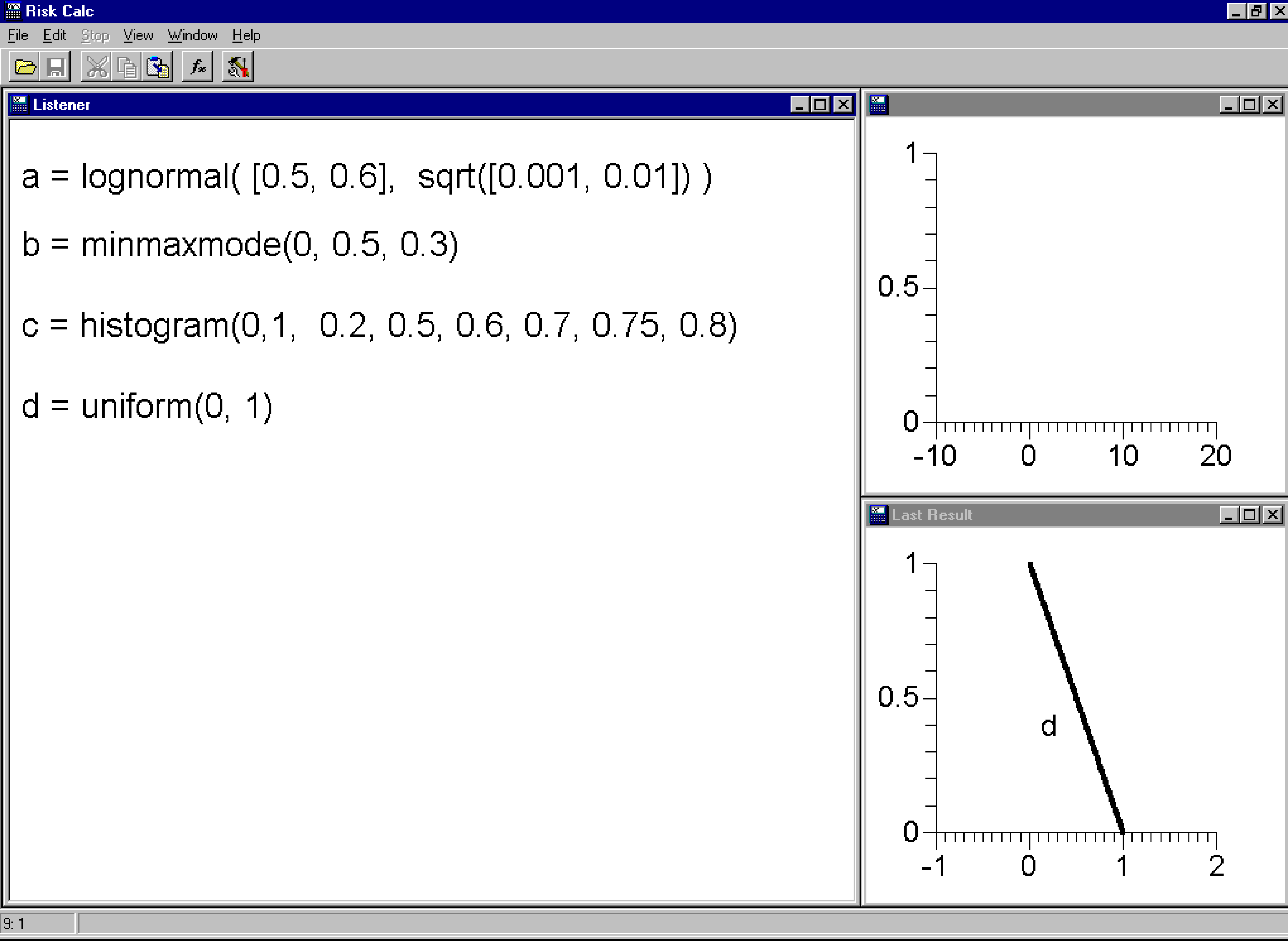

Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
b
-0.1
0.1
0.3
0.5
5: 1



Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
b
-0.1 0.1 0.3 0.5
6: 51


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
c
-1 0 1 2
7: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
c
-1
0
1
2
8: 18


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
d
-1
0
1
2
9: 1

Risk Calc

File Edit Stop View Window Help

Listener

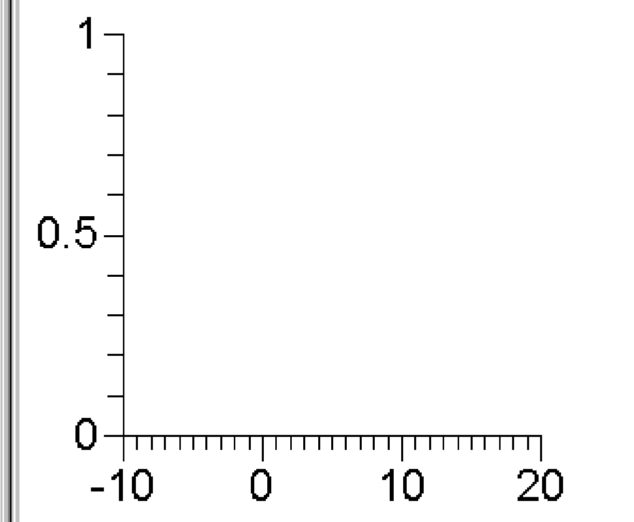
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))

b = minmaxmode(0, 0.5, 0.3)

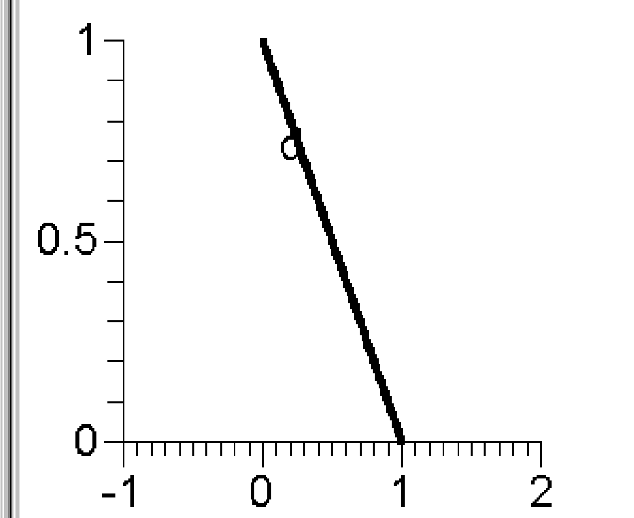
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)

d = uniform(0, 1)

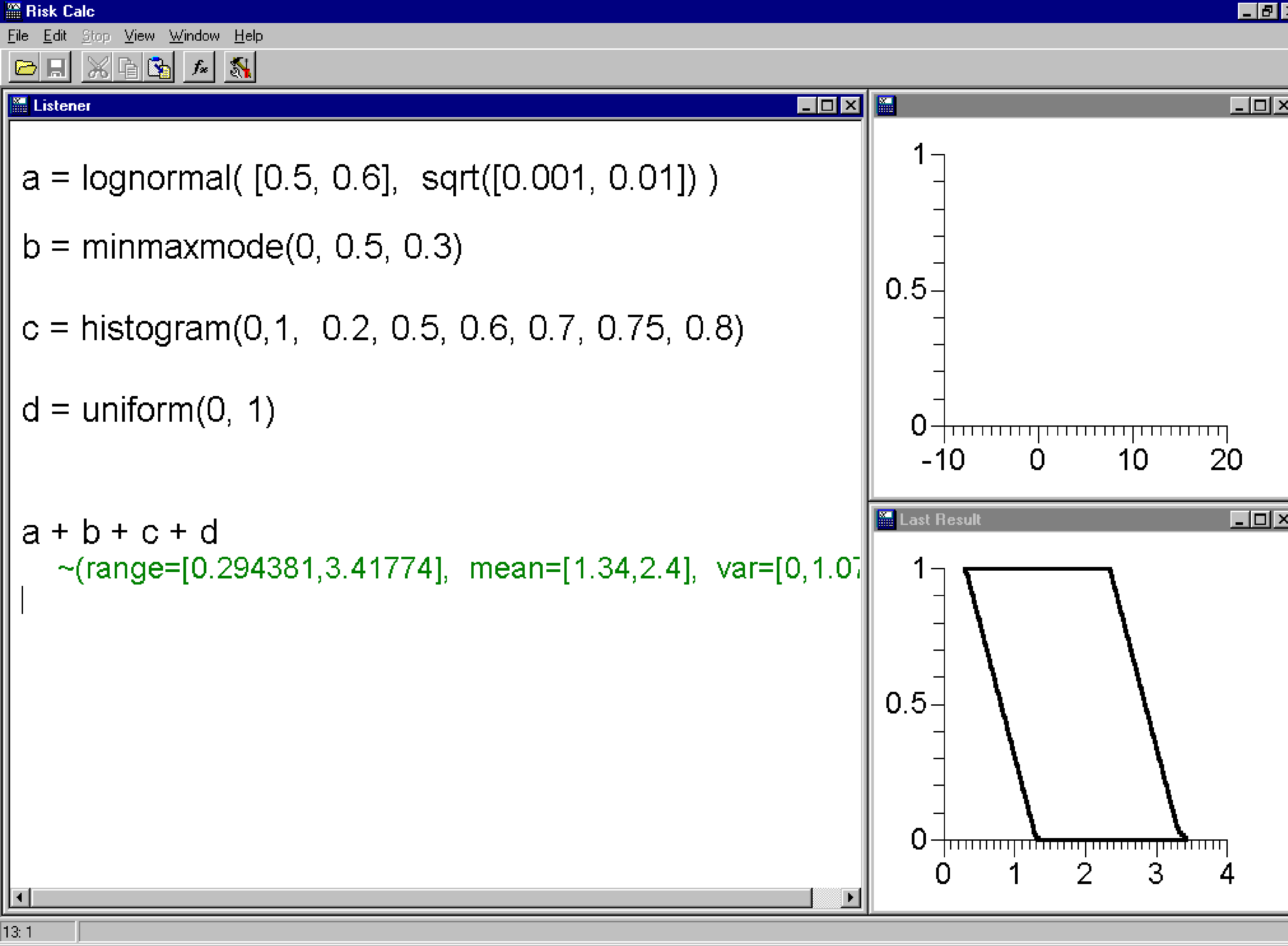
a + b + c + d

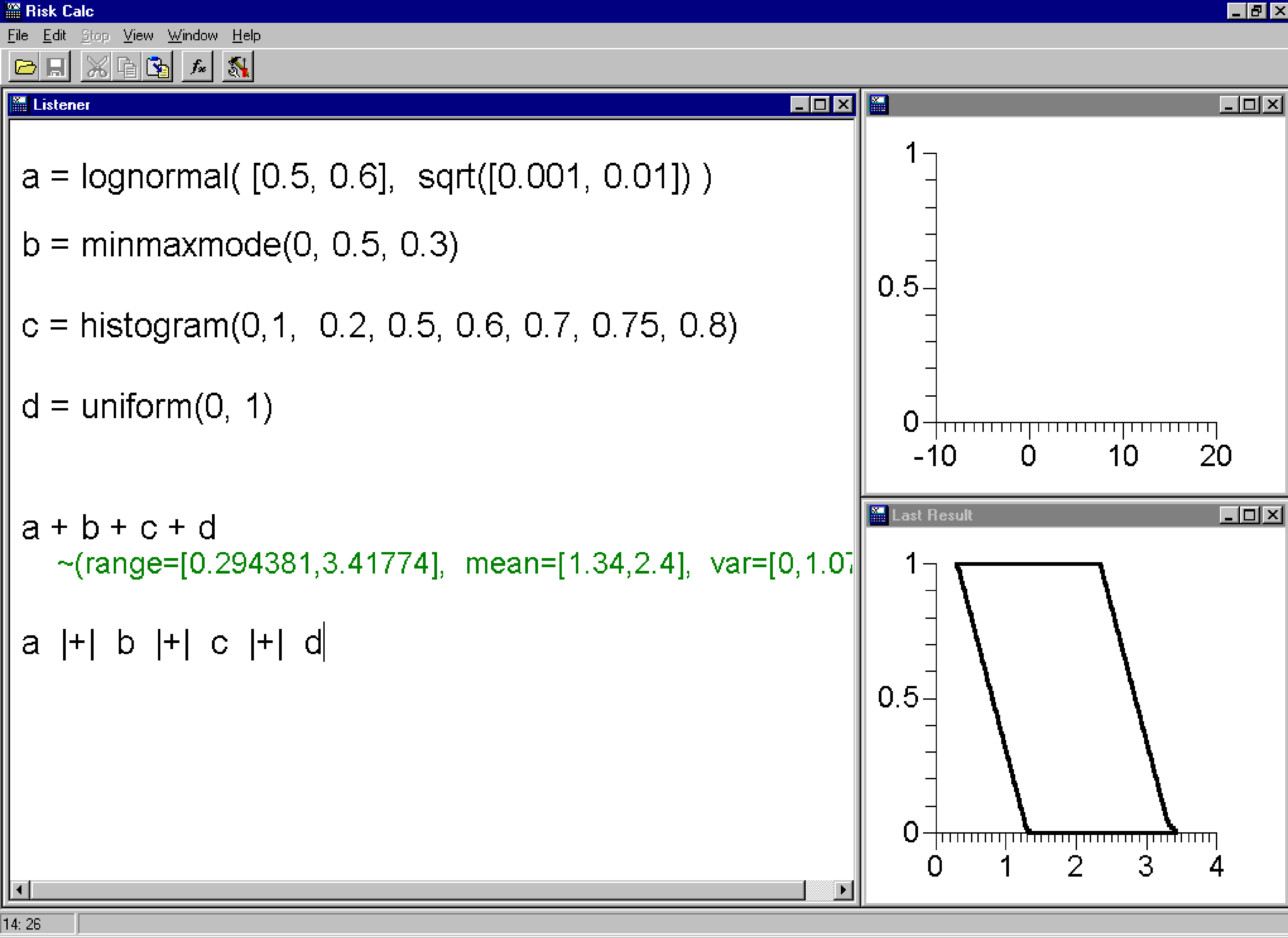


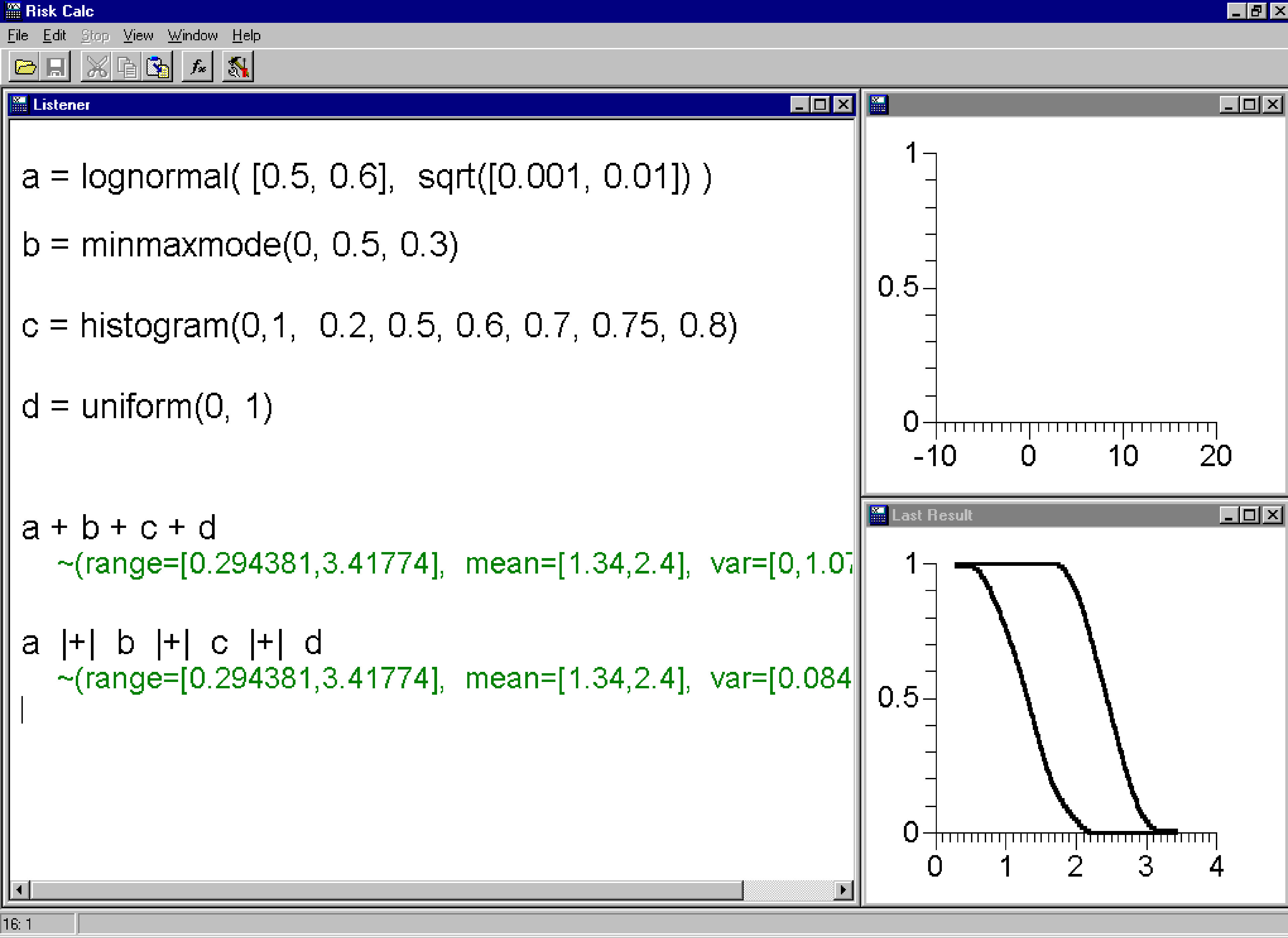
Last Result

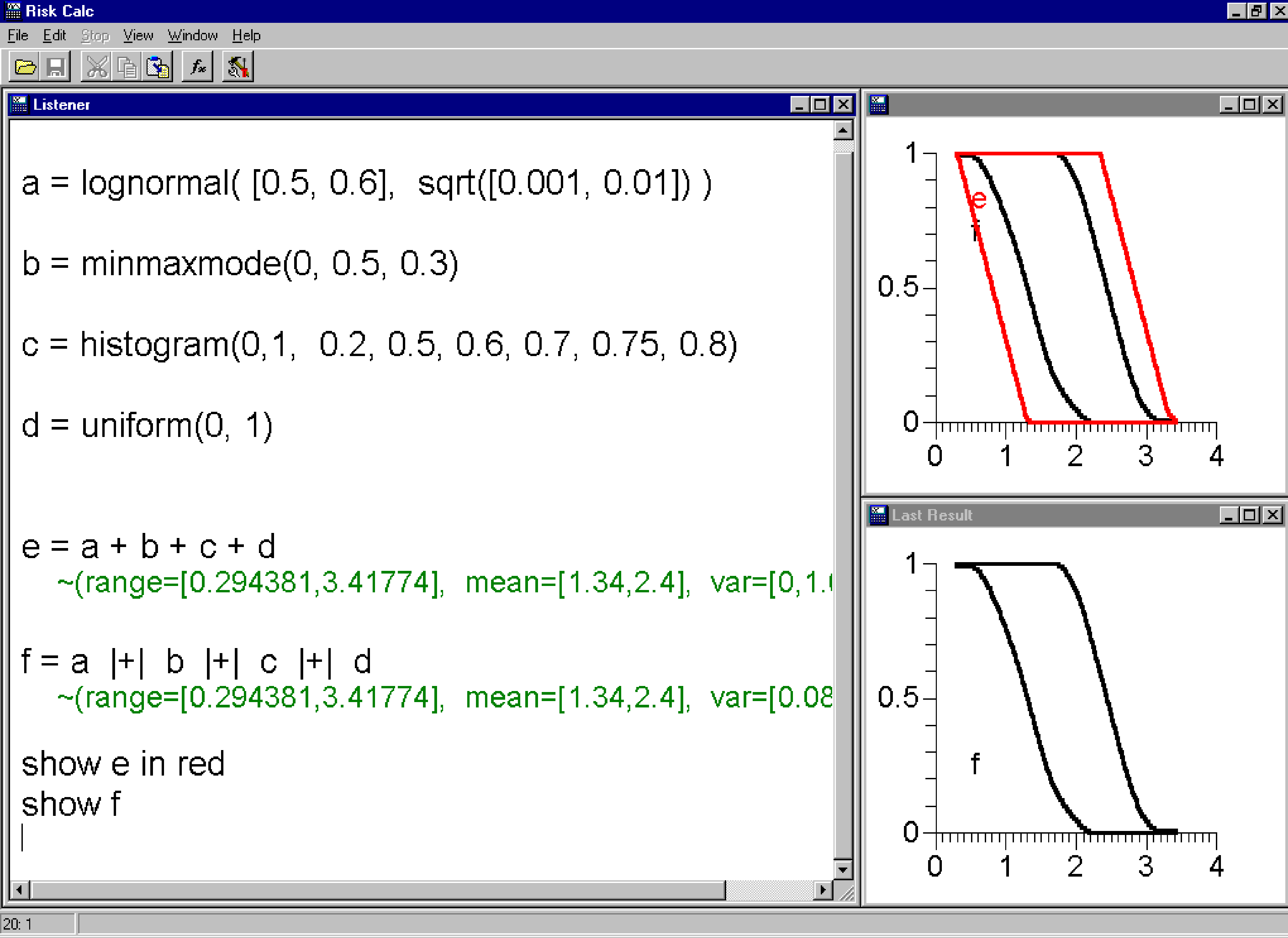


11: 14


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.07
Last Result
13: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.07
a |+| b |+| c |+| d
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
0
1
2
3
4
14: 26


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.07
a |+| b |+| c |+| d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0.084
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
0
1
2
3
4
16: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
e = a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.
f = a |+| b |+| c |+| d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0.08
show e in red
show f
20: 1
e
f
1
0.5
0
0 1 2 3 4
Last Result
f
1
0.5
0
0 1 2 3 4

Diverse applications

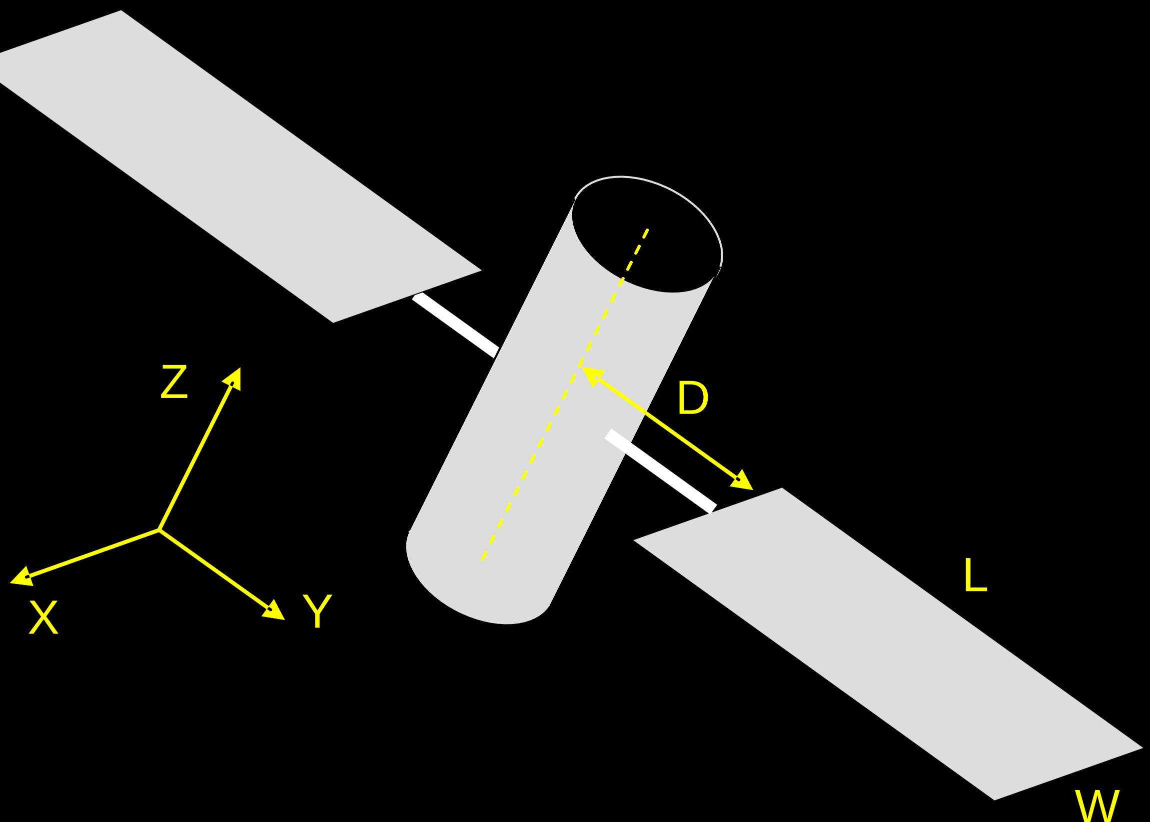
- Superfund risk analyses
- Conservation biology extinction/reintroduction
- Occupational exposure assessment
- Food safety
- Chemostat dynamics
- Global climate change forecasts
- Safety of engineered systems
- Engineering design

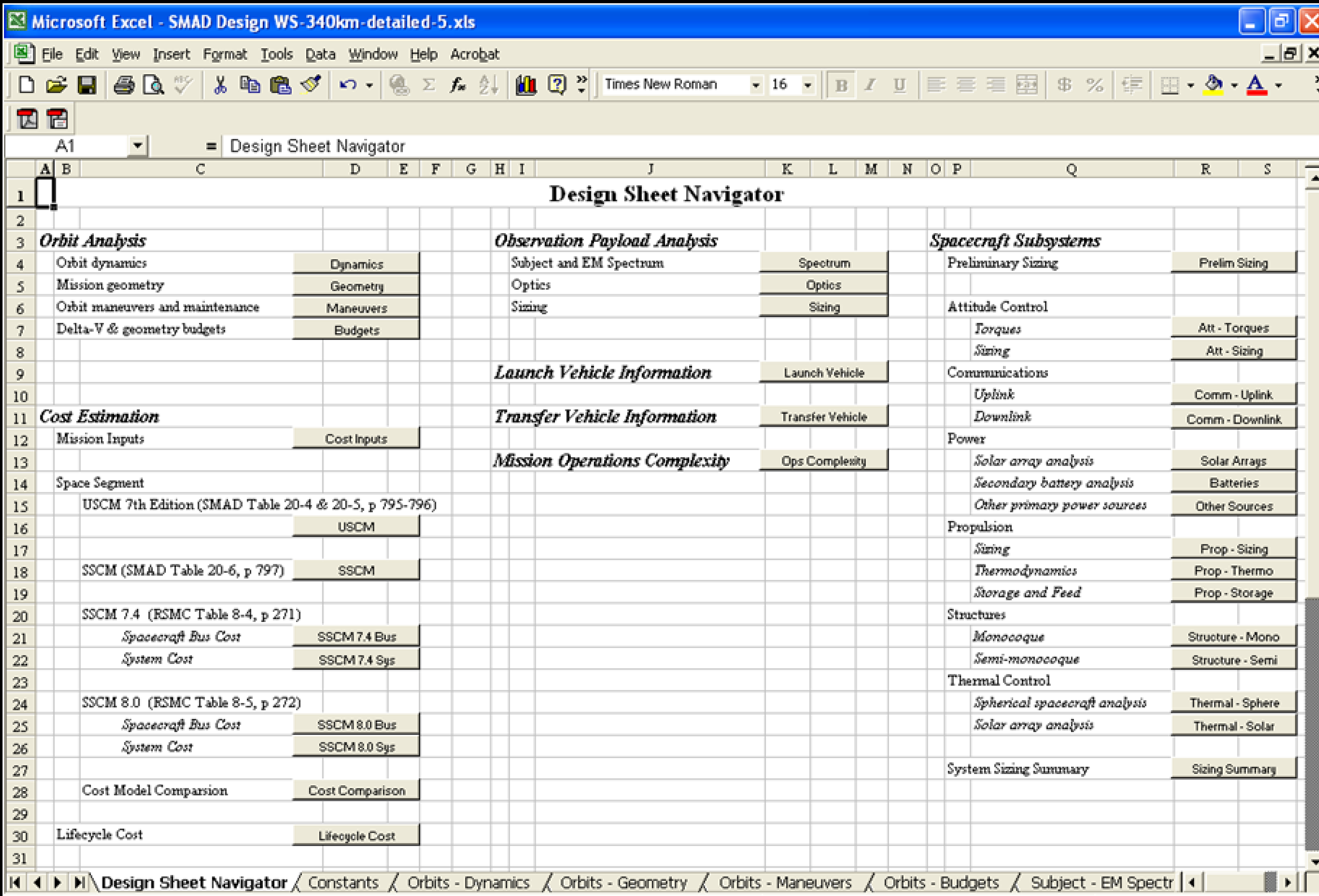
Case study:

Spacecraft design under mission uncertainty

Mission

Deploy satellite carrying a large optical sensor





Wertz and Larson (1999) *Space Mission Analysis and Design (SMAD)*. Kluwer.

Microsoft Excel - SMAD Design WS-340km-detailed-5.xls

Return to Navigator

Attitude Control - Torque Estimates

(All information on this sheet is contained in the block from Cell A1 to Cell Q27)

Orbit characteristics			
Altitude		340.000	km
Satellite velocity		7.703	km/s
Atmospheric density		1.983E-11	kg/m^3
Sun incidence angle		0.00	deg
Maximum deviation from local vertical		10.00	deg
Spacecraft characteristics			
Largest moment of inertia	*7315.000*	7315.000	kg-m^2
Smallest moment of inertia	*4655.000*	4655.000	kg-m^2
Projected surface area	*14.063*	14.063	m^2
Moment arm for solar radiation torques		0.250	m
Moment arm for aerodynamic torques		0.250	m
Drag coefficient		3.13	
Surface reflectivity		0.60	
Residual dipole		1.00	A-m^2
Slew characteristics			
Maximum slewing angle	*38.00*	38.00	deg
Minimum maneuver time	*760.00*	760.00	sec

Environmental torques		
Gravity gradient	1.794E-03	N-m
Solar radiation	2.565E-05	N-m
Magnetic	5.250E-05	N-m
Aerodynamic	6.473E-03	N-m
Total (RSS)	6.717E-03	N-m
Slewing torque	3.360E-02	N-m

Obs Payload - Optics / Obs Payload - Sizing / SC Bus - Preliminary Sizing / **Attitude Control - Torques** / Attitude Control - Sizing / Comm

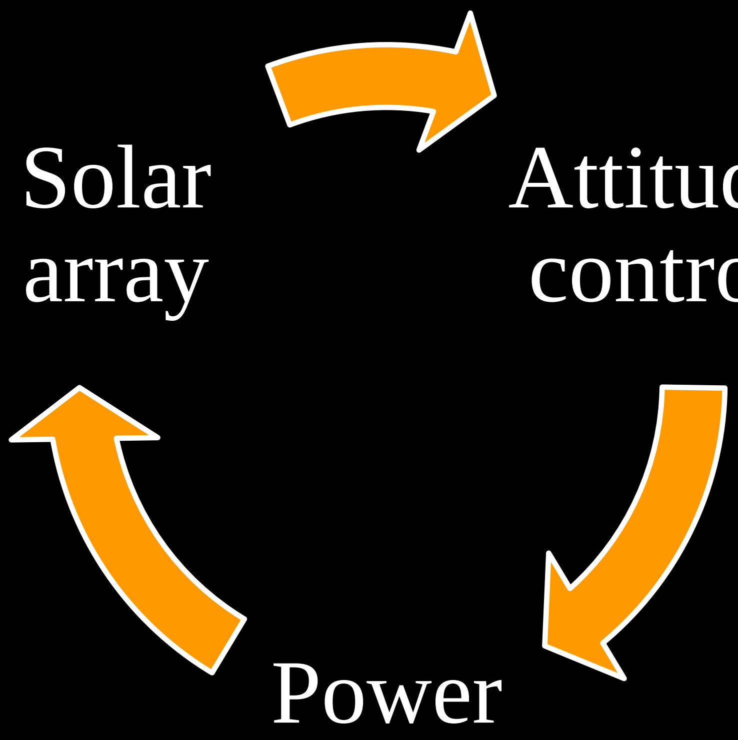
Typical subsystems

- Attitude control
- Command data systems
- Configuration
- Cost
- Ground systems
- Instruments
- Mission design
- Power

- Program management
- Propulsion
- Science
- Solar array
- Systems engineering
- Telecommunications – System
- Telecommunications – Hardware
- Thermal control

Demonstrations

- Calculations within a single subsystem (ACS)
- Calculations within linked subsystems



Attitude control subsystem (ACS)

- 3 reaction wheels
- Design problem: solve for *h*
  - Required angular momentum
  - Needed to choose reaction wheels
- Mission constraints
  - $\Delta t_{\mathrm{orbit}}$ = 1/4 orbit time
  - θ_{slew} = max slew angle
  - Δt_{slew} = min maneuver time
- Inputs from other subsystems
  - I, $I_{\max}$, $I_{\min}$ = inertial moment
  - Depend on solar panel size, which depends on power needed, so on *h*

$$h = \tau_{\mathrm{tot}} \times \Delta t_{\mathrm{orbit}}$$

$$\tau_{tot} = \tau_{slew} + \tau_{dist}$$

$$\tau_{\mathrm{slew}} = \frac{4\theta_{\mathrm{slew}}}{\Delta t_{\mathrm{slew}}^{2}} I$$

$$\tau_{dist} = \tau_{g} + \tau_{sp} + \tau_{m} + \tau_{a}$$

$$\tau_{\mathrm{g}} = \frac{3\mu}{2(R_{\mathrm{E}} + H)^3} \left| I_{\max} - I_{\min} \right| \sin(2\theta)$$

$$\tau_{\mathrm{sp}} = L_{\mathrm{sp}} \frac{F_{\mathrm{s}}}{c} A_{\mathrm{s}} (1 + q) \cos(i)$$

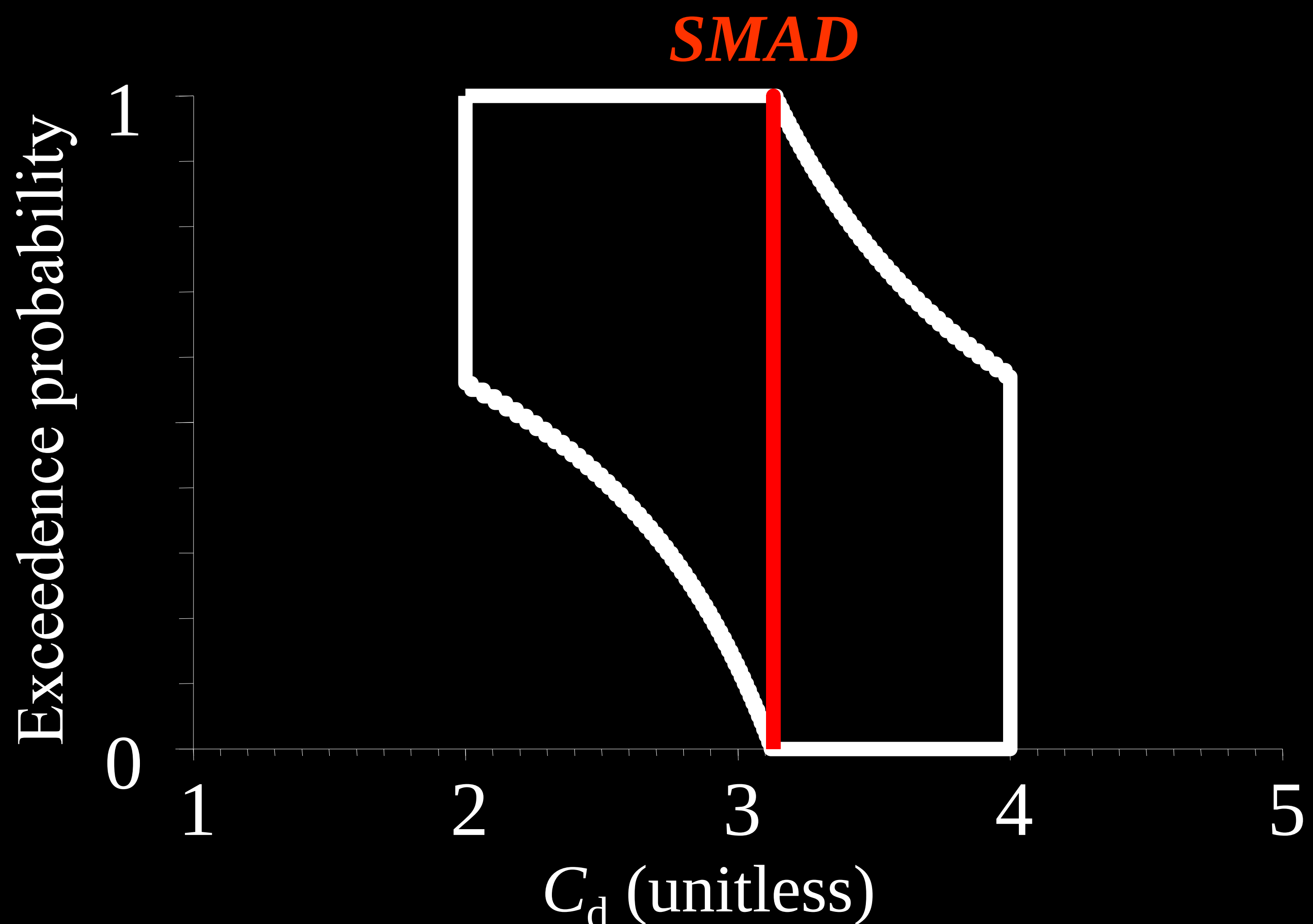
$$\tau_{\mathrm{m}} = \frac{2MD}{(R_{\mathrm{E}} + H)^3}$$

$$\tau_{\mathrm{a}} = \frac{1}{2} L_{\mathrm{a}} \rho C_{\mathrm{d}} A V^2$$

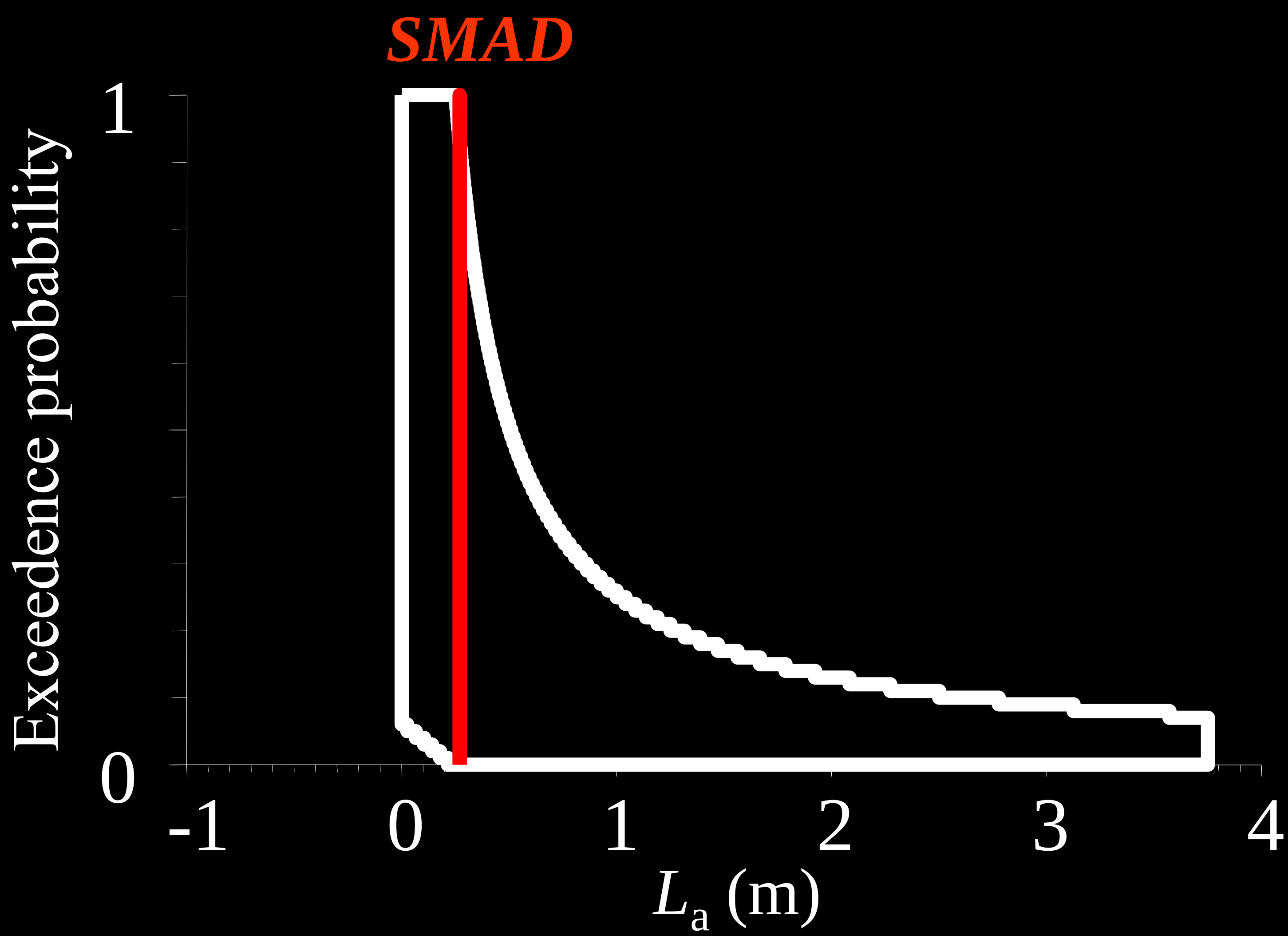
Attitude control input variables

Symbol	Unit	Variable	Type	Value	*SMAD*
C_d	unitless	Drag coefficient	p-box	range=[2,4] mean=3.13	3.13
L_a	m	Aerodynamic drag torque moment	p-box	range=[0,3.75] mean=0.25	0.25
L_{sp}	m	Solar radiation torque moment	p-box	range=[0,3.75] mean=[0.25]	0.25
D_r	A m^2	Residual dipole	interval	[0,1]	1
i	degrees	Sun incidence angle	interval	[0,90]	0
ρ	kg m^3	Atmospheric density	interval	[3.96e-12, 9.9e-11]	1.98e-11
θ	degrees	Major moment axis deviation from nadir	interval	[10,19]	10
q	unitless	Surface reflectivity	interval	[0.1,0.99]	0.6
I_{min}	kg m^2	Minimum moment of inertia	interval	[4655]	4655
I_{max}	kg m^2	Maximum moment of inertia	interval	[7315]	7315
μ	m^3 s^{-2}	Earth gravity constant	point	3.98e14	3.98e14
A	m^2	Area in the direction of flight	point	3.75^2	3.75^2
RE	km	Earth radius	point	6378.14	6378.14
H	km	Orbit altitude	point	340	340
F_s	W m^{-2}	Average solar flux	point	1367	1367
θ_{slew}	degrees	Maximum slewing angle	point	38	38
c	m s^{-1}	Light speed	point	2.9979e8	2.9979e8
M	A m^2	Earth magnetic moment	point	7.96e22	7.96e22
Δt_{slew}	s	Minimum maneuver time	point	760	760
A_s	m^2	Area reflecting solar radiation	point	3.75^2	3.75^2
Δt_{orbit}	s	Quarter orbit period	point	1370	1370

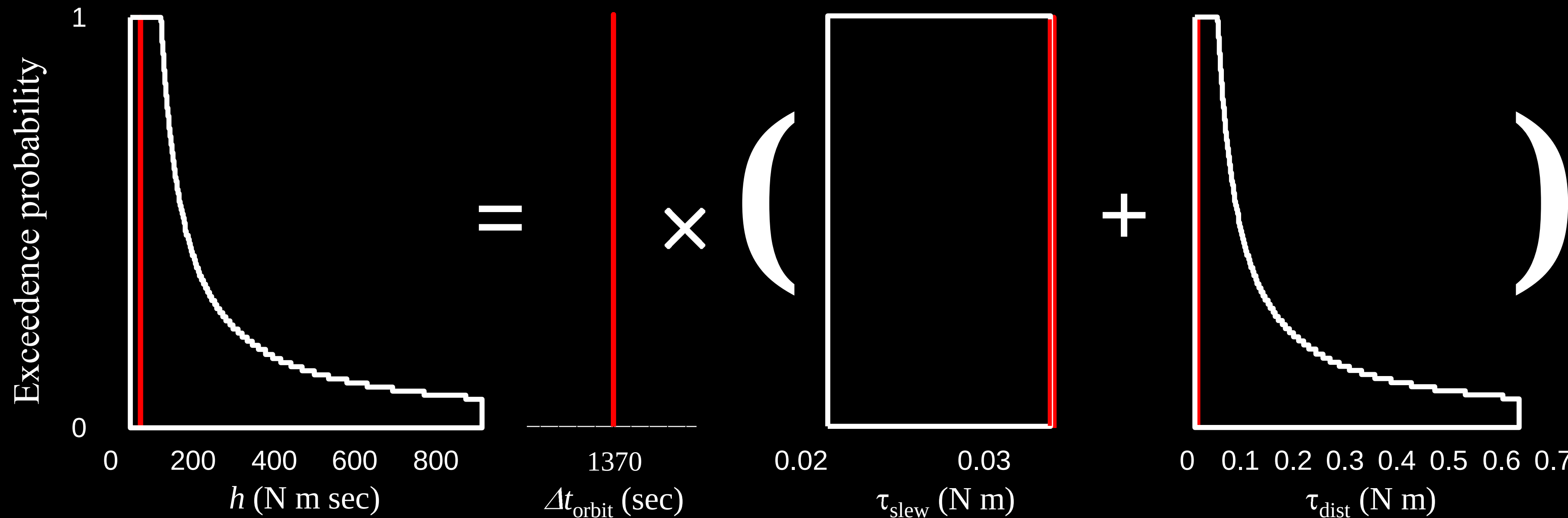
Coefficient of drag, C_d



Aerodynamic drag torque moment, L_a

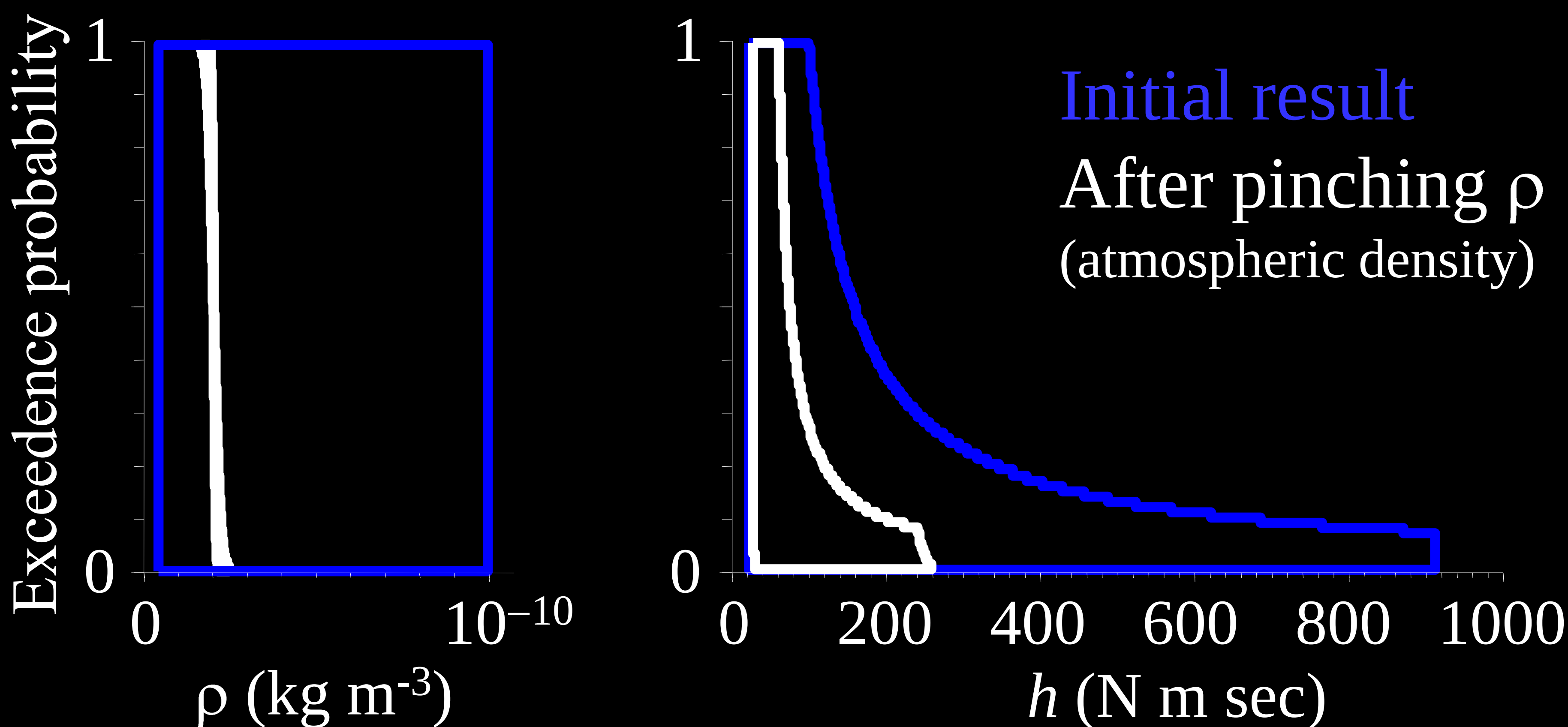


Required angular momentum *h*

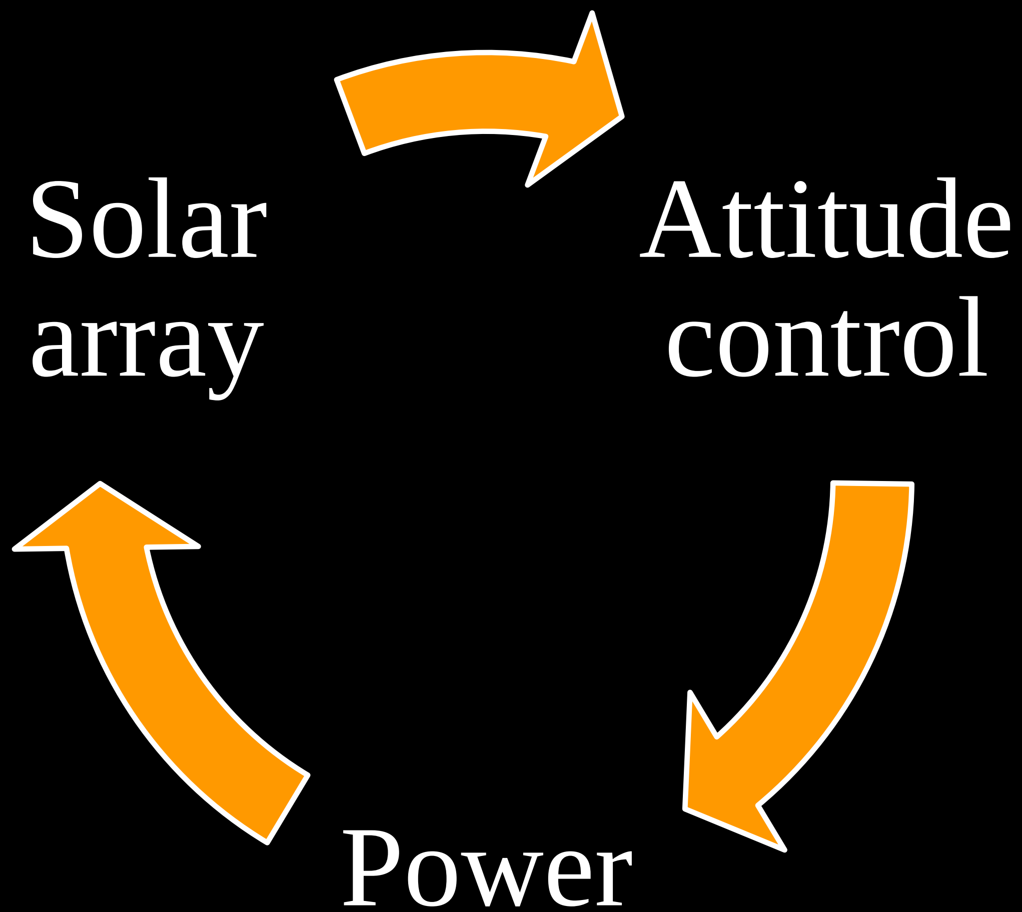


$$h = \Delta t_{orbit} \times (\tau_{slew} + \tau_{dist})$$

Value of information: pinching ρ



Three linked subsystems



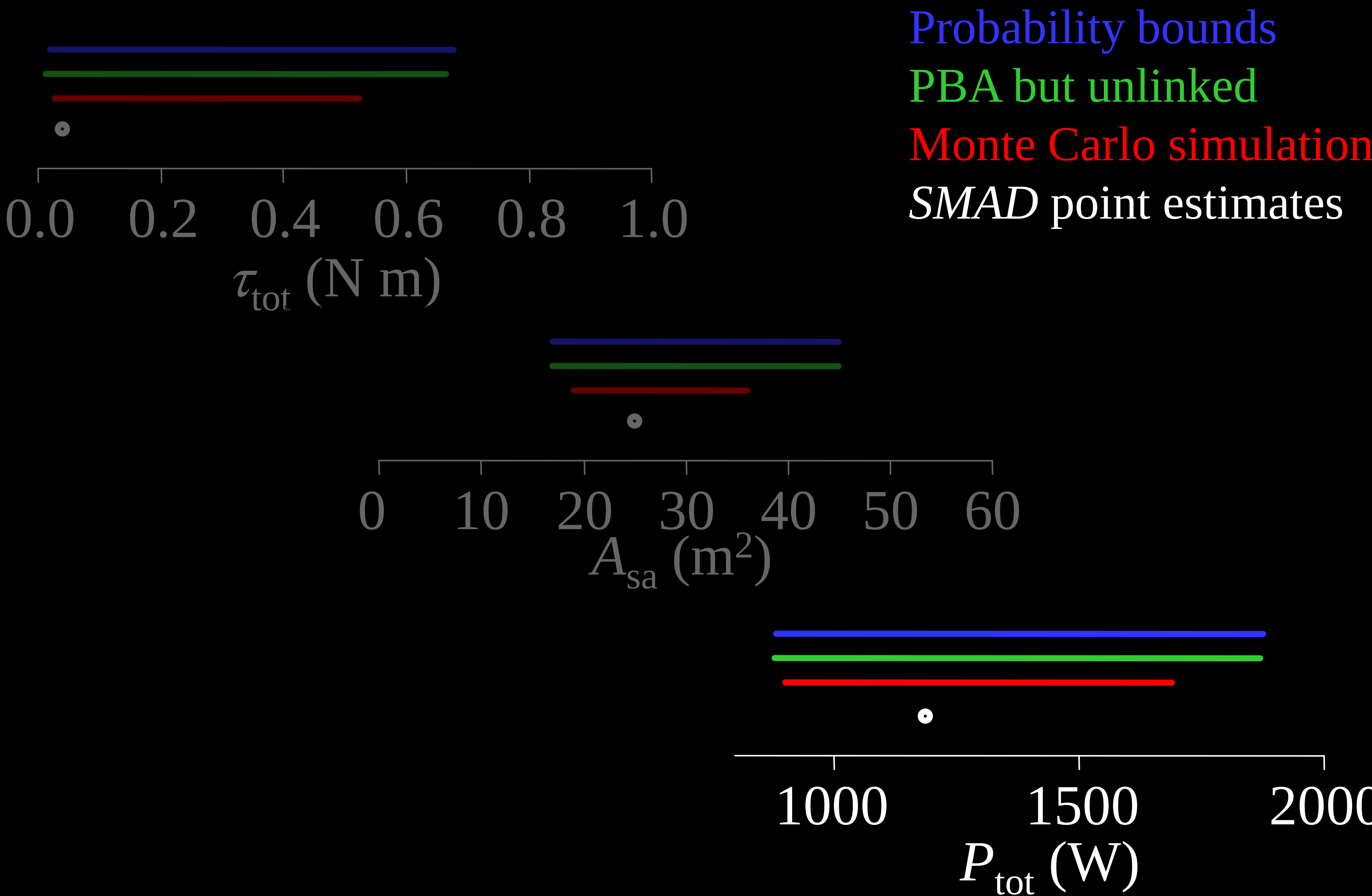
Variables passed iteratively

- Minimum moment of inertia I_{min}
- Maximum moment of inertia I_{max}
- Total torque τ_{tot}
- Total power P_{tot}
- Solar panel area A_{sa}

Analysis of calculations

- Need to check that original *SMAD* values and all Monte Carlo simulations are enclosed by p-boxes
- Need to ensure iteration through links doesn't cause runaway uncertainty growth (or reduction)
- Four parallel analyses
  - *SMAD*'s point estimates
  - Monte Carlo simulation
  - P-boxes but without linkage among subsystems
  - P-boxes with fully linked subsystems

Supports of results



Case study findings

- Different answers are consistent
  - Point estimates match the *SMAD* results
  - P-boxes span the points and the Monte Carlo intervals
- Calculations workable
  - No runaway inflation (or loss) of uncertainty
  - Easier than with Monte Carlo
- Practical and interesting results
  - Uncertainty can affect engineering decisions
  - Reducing uncertainty about ρ (by picking a launch date) strongly reduces design uncertainty

For more information, consult SBIR project report to NASA, July 2009

Accounting for epistemic and aleatory uncertainty in early system design

NASA SBIR Phase 2 Final Report

Applied Biomathematics
100 North Country Road
Setauket, NY 11733

Phone: 1-631-751-4350
Fax: 1-631-751-3435
Email: scott@ramas.com
www.ramas.com

Order Number: NNL07AA06C

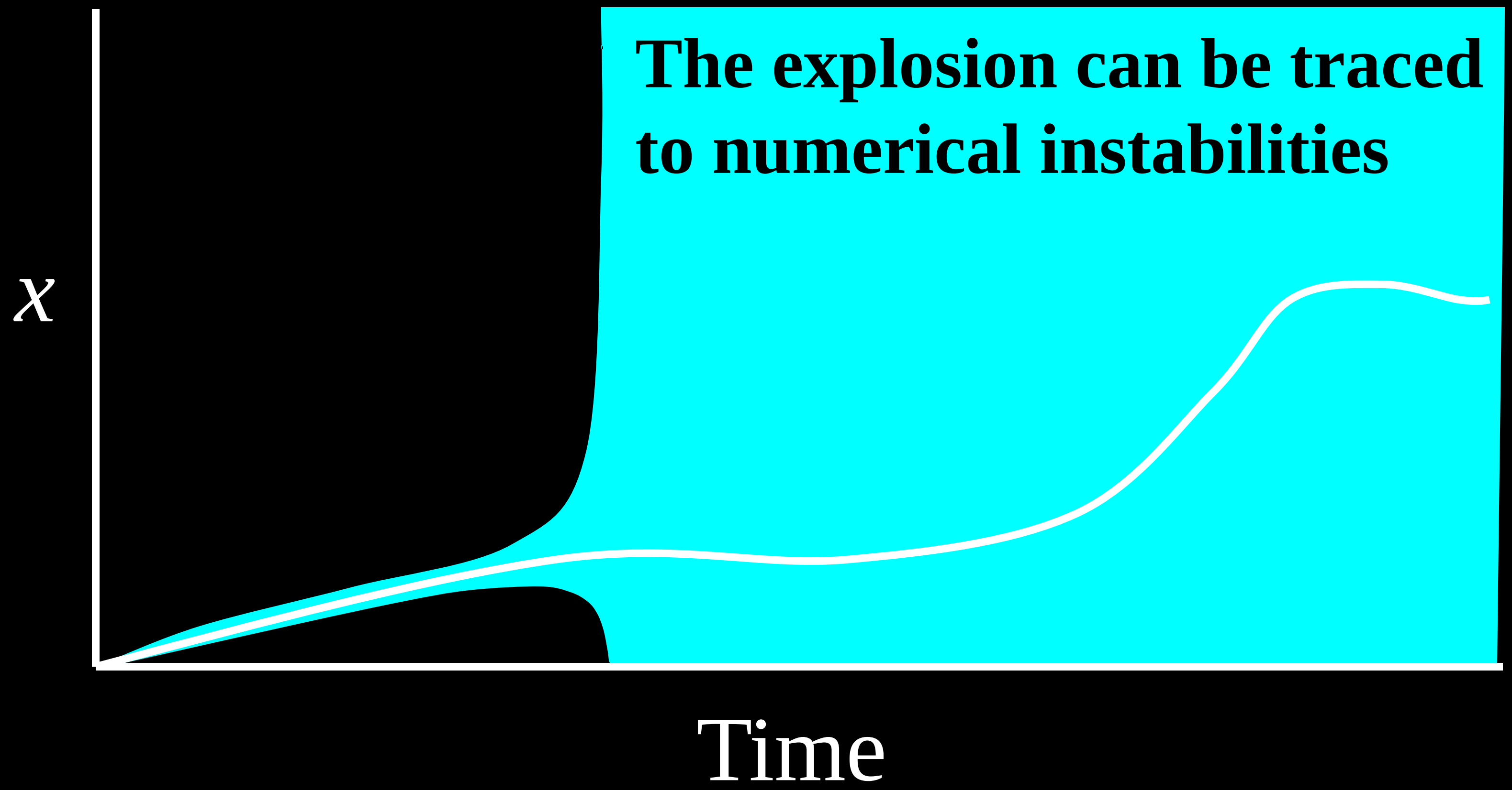
July 2009

Uses for probability bounds analysis

- Uncertainty propagation
- Risk assessment
- Sensitivity analysis (for control and study)
- Reliability theory
- Engineering design ✓
- Validation
- Decision theory
- Regulatory compliance
- Finite element modeling
- Differential equations ✓

Differential equations

Uncertainty usually explodes



Uncertainty

- Artifactual uncertainty
  - Too few polynomial terms
  - Numerical instability
  - Can be reduced by a better analysis

- Authentic uncertainty
  - Genuine unpredictability due to input uncertainty
  - Cannot be reduced by a better analysis

Only by more information, data or assumptions

Uncertainty propagation

- We *want* the prediction to ‘break down’ if that’s what should happen
- But we don’t want artifactual uncertainty
  - Numerical instabilities
  - Wrapping effect
  - Dependence problem
  - Repeated parameters

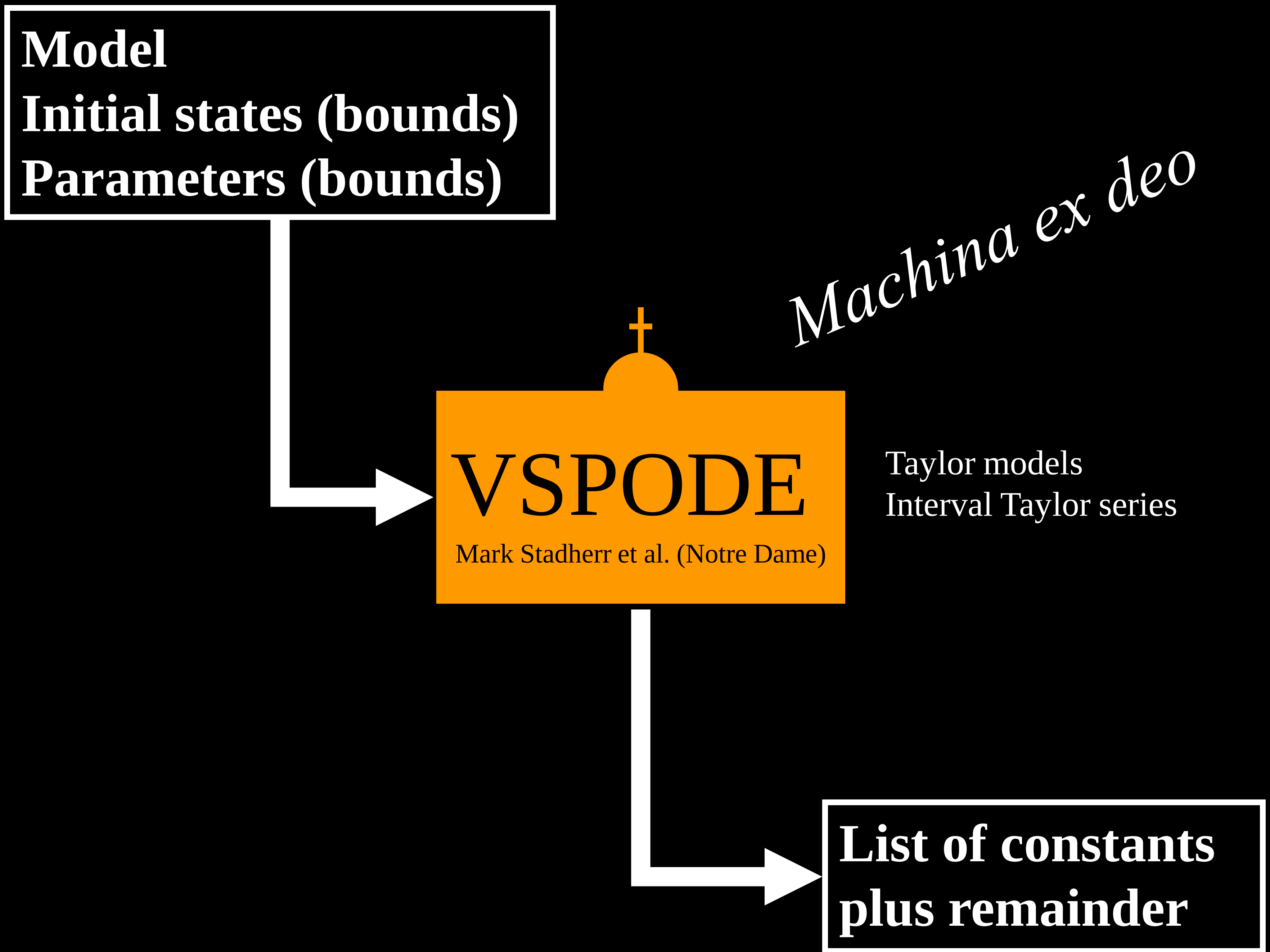
Problem

- Nonlinear ordinary differential equation (ODE)

$$\mathrm{d}\boldsymbol{x}/\mathrm{d}t = f(\boldsymbol{x}, \theta)$$

with uncertain θ and uncertain initial state x_0

- Information about θ and x_0 comes as
  - Interval ranges
  - Probability distributions
  - Probability boxes


Model
Initial states (bounds)
Parameters (bounds)
Machina ex deo
VSPODE
Mark Stadherr et al. (Notre Dame)
Taylor models
Interval Taylor series
List of constants
plus remainder

Example ODE

$$dx_1/dt = \theta_1\, x_1(1 - x_2)$$
$$dx_2/dt = \theta_2\, x_2(x_1 - 1)$$

What are the states at $t = 10$?

$$x_0 = (1.2, 1.1)^T$$
$$\theta_1 \in [2.99, 3.01]$$
$$\theta_2 \in [0.99, 1.01]$$

VSPODE

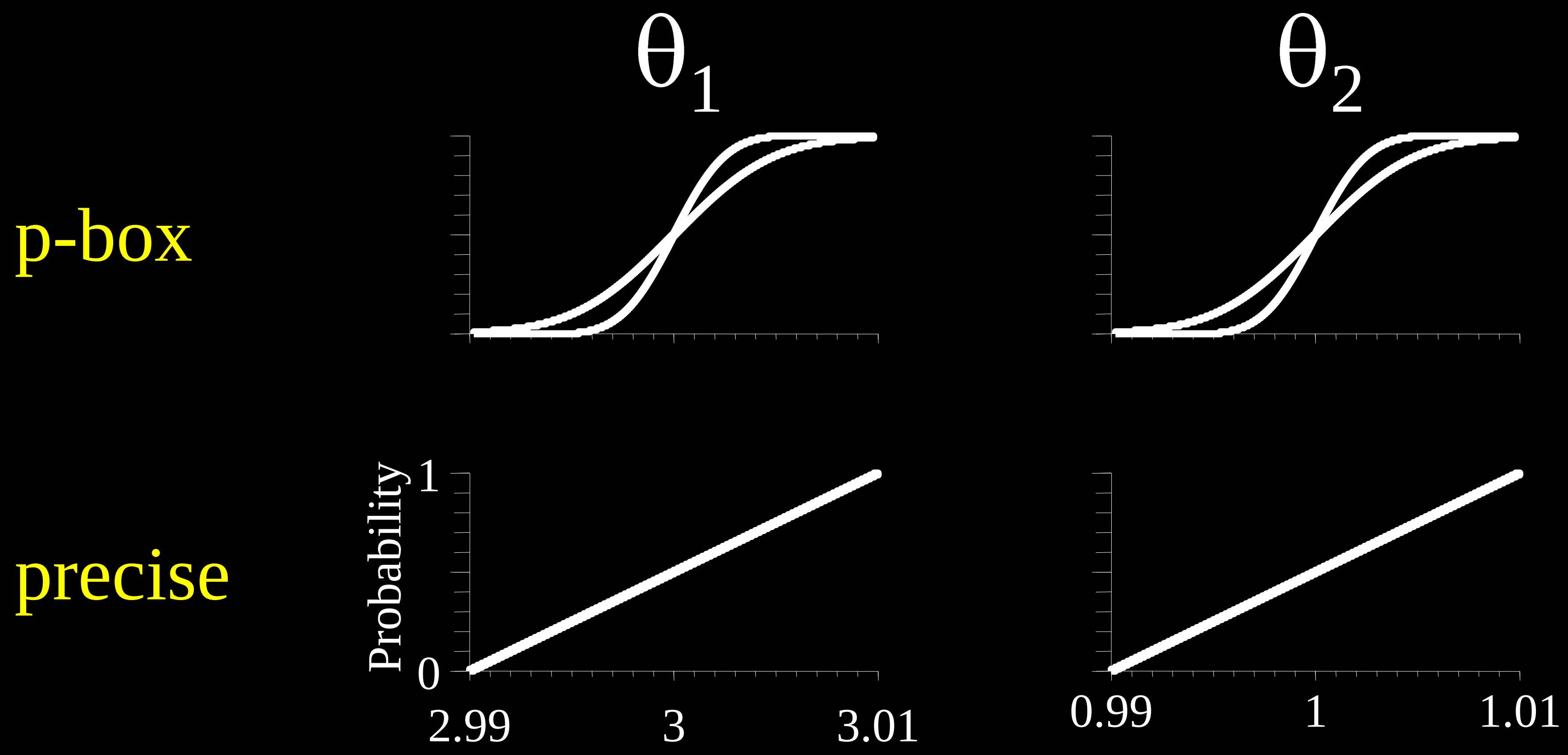
- Constant step size $h = 0.1$, Order of Taylor model $q = 5$,
- Order of interval Taylor series $k = 17$, QR factorization

VSPODE tells how to compute x_1

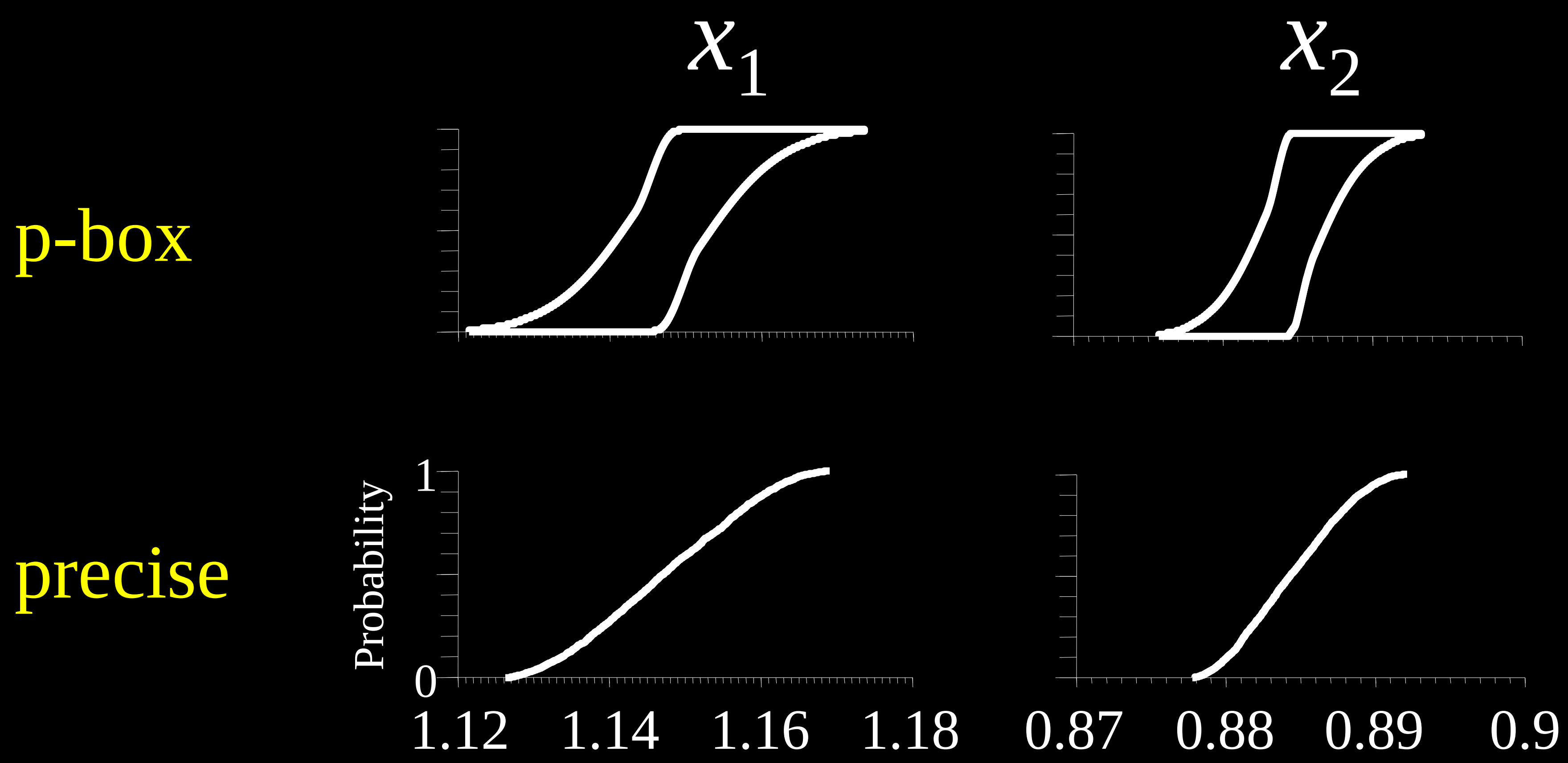
$$
\begin{aligned}
&1.916037656181642 \times \theta_1^{\,0} \times \theta_2^{\,1} \;+\; 0.689979149231081 \times \theta_1^{\,1} \times \theta_2^{\,0} \;+\\
&-4.690741189299572 \times \theta_1^{\,0} \times \theta_2^{\,2} \;+\; -2.275734193378134 \times \theta_1^{\,1} \times \theta_2^{\,1} \;+\\
&-0.450416914564394 \times \theta_1^{\,2} \times \theta_2^{\,0} \;+\; -29.788252573360062 \times \theta_1^{\,0} \times \theta_2^{\,3} \;+\\
&-35.200757076497972 \times \theta_1^{\,1} \times \theta_2^{\,2} \;+\; -12.401600707197074 \times \theta_1^{\,2} \times \theta_2^{\,1} \;+\\
&-1.349694561113611 \times \theta_1^{\,3} \times \theta_2^{\,0} \;+\; 6.062509834147210 \times \theta_1^{\,0} \times \theta_2^{\,4} \;+\\
&-29.503128650484253 \times \theta_1^{\,1} \times \theta_2^{\,3} \;+\; -25.744336555602068 \times \theta_1^{\,2} \times \theta_2^{\,2} \;+\\
&-5.563350070358247 \times \theta_1^{\,3} \times \theta_2^{\,1} \;+\; -0.222000132892585 \times \theta_1^{\,4} \times \theta_2^{\,0} \;+\\
&218.607042326120308 \times \theta_1^{\,0} \times \theta_2^{\,5} \;+\; 390.260443722081675 \times \theta_1^{\,1} \times \theta_2^{\,4} \;+\\
&256.315067368131281 \times \theta_1^{\,2} \times \theta_2^{\,3} \;+\; 86.029720297509172 \times \theta_1^{\,3} \times \theta_2^{\,2} \;+\\
&15.322357274648443 \times \theta_1^{\,4} \times \theta_2^{\,1} \;+\; 1.094676837431721 \times \theta_1^{\,5} \times \theta_2^{\,0} \;+\\
&[\; \underline{1.14775}37620811058,\; \underline{1.14775}39164945061 \;]
\end{aligned}
$$

where θ's are centered forms of the parameters; $\theta_1 = \mathrm{\theta}_1 - 3$, $\theta_2 = \mathrm{\theta}_2 - 1$

Input p-boxes



Results


x_1
x_2
p-box
precise
Probability
1
0
1.12
1.14
1.16
1.18
0.87
0.88
0.89
0.9

Still repeated uncertainties

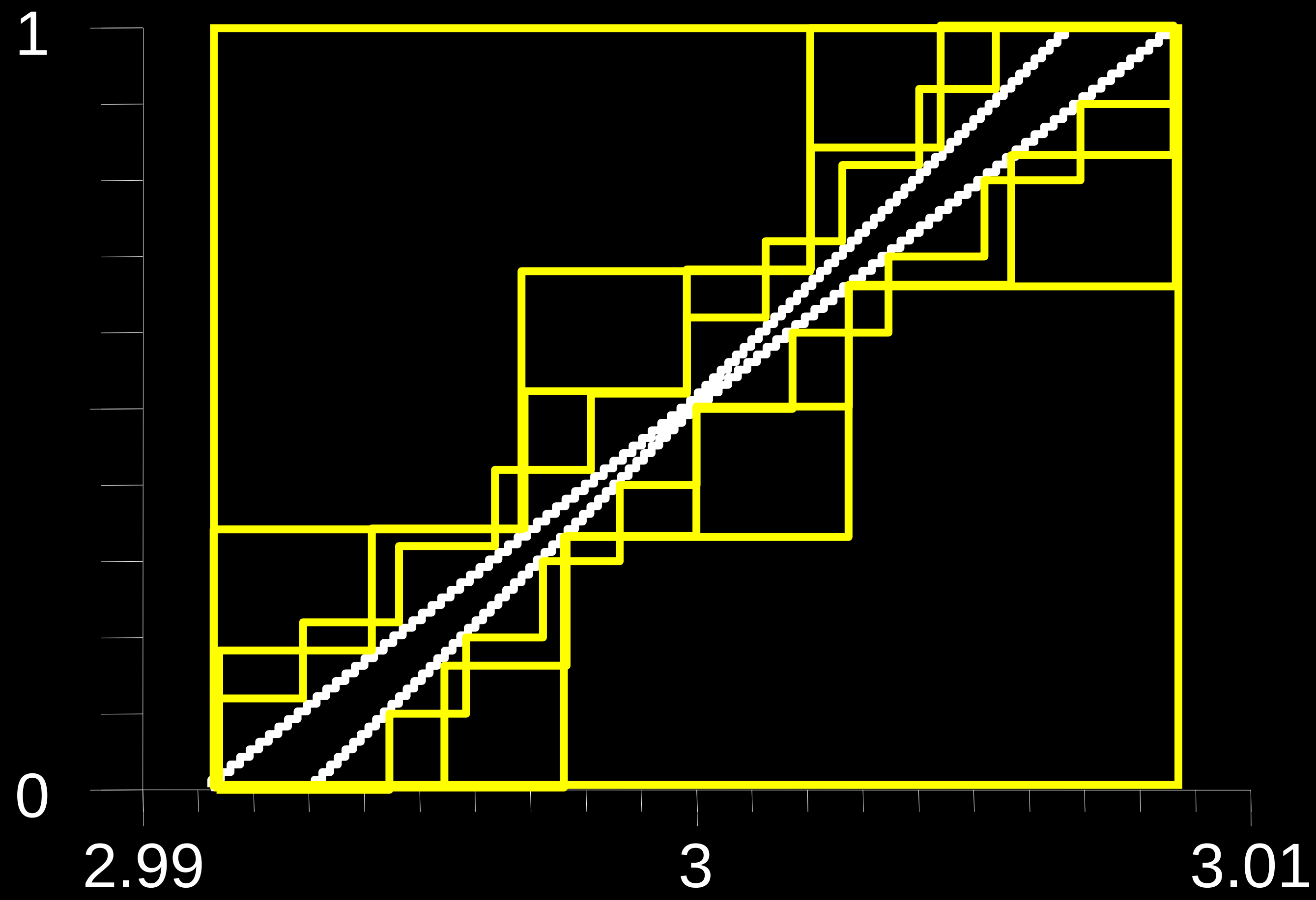
$$
\begin{array}{rr}
1.916037656181642 \times \theta_1^{0} \times \theta_2^{1} \; + & 0.689979149231081 \times \theta_1^{1} \times \theta_2^{0} \; + \\
-4.690741189299572 \times \theta_1^{0} \times \theta_2^{2} \; + & -2.275734193378134 \times \theta_1^{1} \times \theta_2^{1} \; + \\
-0.450416914564394 \times \theta_1^{2} \times \theta_2^{0} \; + & -29.788252573360062 \times \theta_1^{0} \times \theta_2^{3} \; + \\
-35.200757076497972 \times \theta_1^{1} \times \theta_2^{2} \; + & -12.401600707197074 \times \theta_1^{2} \times \theta_2^{1} \; + \\
-1.349694561113611 \times \theta_1^{3} \times \theta_2^{0} \; + & 6.062509834147210 \times \theta_1^{0} \times \theta_2^{4} \; + \\
-29.503128650484253 \times \theta_1^{1} \times \theta_2^{3} \; + & -25.744336555602068 \times \theta_1^{2} \times \theta_2^{2} \; + \\
-5.563350070358247 \times \theta_1^{3} \times \theta_2^{1} \; + & -0.222000132892585 \times \theta_1^{4} \times \theta_2^{0} \; + \\
218.607042326120308 \times \theta_1^{0} \times \theta_2^{5} \; + & 390.260443722081675 \times \theta_1^{1} \times \theta_2^{4} \; + \\
256.315067368131281 \times \theta_1^{2} \times \theta_2^{3} \; + & 86.029720297509172 \times \theta_1^{3} \times \theta_2^{2} \; + \\
15.322357274648443 \times \theta_1^{4} \times \theta_2^{1} \; + & 1.094676837431721 \times \theta_1^{5} \times \theta_2^{0} \; +
\end{array}
$$

[1.1477537620811058, 1.1477539164945061]

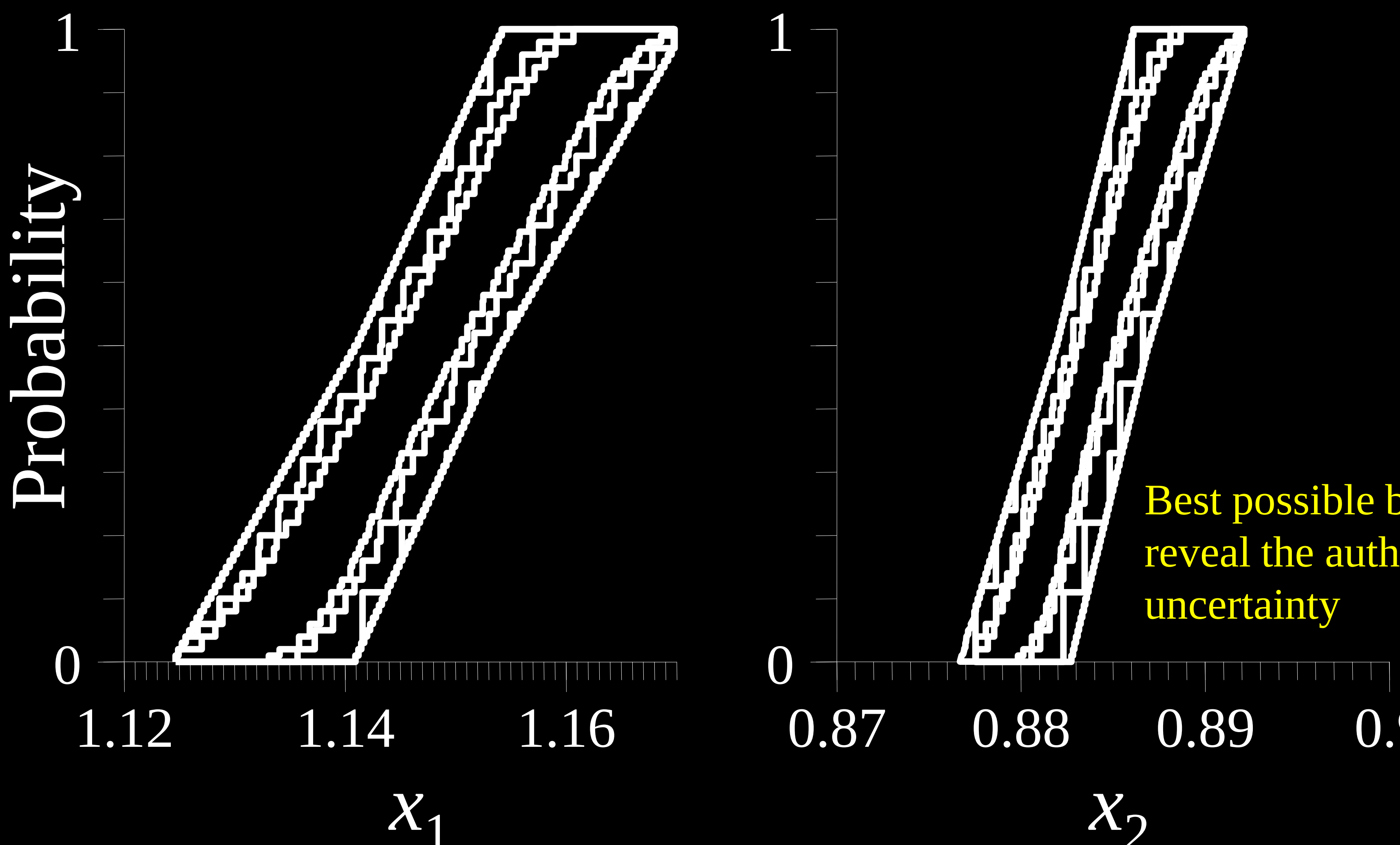
Subinterval reconstitution

- Subinterval reconstitution (SIR)
  - Partition the inputs into subintervals
  - Apply the function to each subinterval
  - Form the union of the results
- Still rigorous, but often tighter
  - The finer the partition, the tighter the union
  - Many strategies for partitioning
- Apply to *each cell* in the Cartesian product

Discretizations



Contraction from SIR



Best possible bounds reveal the authentic uncertainty

Monte Carlo is more limited

- Monte Carlo cannot propagate incertitude
- Monte Carlo *cannot produce validated results*
  - Though can be checked by repeating simulation
- Validated results from distributions can be obtained by modeling inputs with (narrow) p-boxes and applying probability bounds analysis
- Results converge to narrow p-boxes obtained from infinitely many Monte Carlo replications

Results

- Probability bounds analysis with VSPODE are useful for bounding solutions of nonlinear ODEs
- They rigorously propagate uncertainty

about

Initial states

Parameters

in the form of

Intervals

Distributions

P-boxes

Paper in *AIChE Journal* [*American Institute of Chemical Engineers*], February 2011 (on line May 2010)

AIChE

Probability Bounds Analysis for Nonlinear Dynamic Process Models

Joshua A. Enszer and Youdong Lin
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

Scott Ferson
Applied Biomathematics, Setauket, NY 11733

George F. Corliss
Dept. of Electrical and Computer Engineering, Marquette University, Milwaukee, WI 53201

Mark A. Stadtherr
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

DOI 10.1002/aic.12278
Published online May 4, 2010 in Wiley Online Library (wileyonlinelibrary.com).

Dynamic process models frequently involve uncertain parameters and inputs. Propagating these uncertainties rigorously through a mathematical model to determine their effect on system states and outputs is a challenging problem. In this work, we describe a new approach, based on the use of Taylor model methods, for the rigorous propagation of uncertainties through nonlinear systems of ordinary differential equations (ODEs). We concentrate on uncertainties whose distribution is not known precisely, but can be bounded by a probability box (p-box), and show how to use p-boxes in the context of Taylor models. This allows us to obtain p-box representations of the uncertainties in the state variable outputs of a nonlinear ODE model. Examples having two to three uncertain parameters or initial states and focused on reaction process dynamics are used to demonstrate the potential of this approach. Using this method, rigorous probability bounds can be determined at a computational cost that is significantly less than that required by Monte Carlo analysis. © 2010 American Institute of Chemical Engineers *AIChE J*, 57: 404–422, 2011

Keywords: design (process simulation), mathematical modeling, numerical solutions, reactor analysis, bioprocess engineering

Introduction

Systems of ordinary differential equations (ODEs) are the basis for many mathematical models in engineering and science. For example, models of reactor dynamics are based on unsteady-state material and energy balances and thus take the form of a system of first-order ODEs, which typically is nonlinear. Generally, the problem of interest is an initial value problem (IVP), in which an initial state is given and the system then integrated numerically until some final time (time horizon) is reached, thus determining numerical approximations of the final state, as well as of the trajectory followed to reach it.

Often these dynamic models involve uncertainties in parameters and/or initial states. Analysis of the impact of such uncertainties is clearly important in models of process dynamics, as used, for example, in state and parameter estimation[1,2]

Correspondence concerning this article should be addressed to M. A. Stadtherr at markst@nd.edu.
Current Address of Youdong Lin: LINDO Systems, Inc., 1415 North Dayton St., Chicago, IL 60642.

PBA relaxes assumptions

- Everyone makes assumptions, but not all sets of assumptions are equal:

Linear	Normal	Independence
Montonic	Unimodal	Known correlation
Any function	***Any distribution***	***Any dependence***

- PBA doesn't require unwarranted assumptions

Wishful thinking

Analysts often make convenient assumptions that are not really justified:

1. Variables are independent of one another
2. Uniform distributions model gross incertitude
3. Distributions are stationary (unchanging)
4. Distributions are perfectly precisely specified
5. Measurement uncertainty is negligible

You don’t have to think wishfully

A p-box can discharge a false assumption:

1. Don’t have to assume any dependence at all
2. An interval can be a better model of incertitude
3. P-boxes can enclose non-stationary distributions
4. Can handle imprecise specifications
5. Measurement data with plus-minus, censoring

Rigorousness

- "Automatically verified calculations"
- The computations are guaranteed to enclose the true results (so long as the inputs do)
- You can still be wrong, but the *method* won't be the reason if you are

Take-home messages

- Using bounding, you don’t have to pretend you know a lot to get quantitative results

- Probability bounds analysis bridges worst case and probabilistic analyses in a way that’s faithful to both and makes it suitable for use in early design

Acknowledgments

Larry Green, LaRC

Bill Oberkampf, Sandia

Chris Paredis, Georgia Tech

NASA

National Institutes of Health (NIH)

Electric Power Research Institute (EPRI)

Sandia National Laboratories

End

Failure Domain Bounding with Applications to Dynamic Systems

Sean P. Kenny, Ph.D. and Daniel P. Giesy, Ph.D.
Dynamic Systems & Control Branch
NASA Langley Research Center

Luis G. Crespo, Ph.D.
Senior Research Scientist
National Institute of Aerospace

May 4, 2011

Favorite Quote from NSES 2011

“When systems fail, people notice.”

Dr. J Michael Gilmore

Recent NASA History and Context

- **2003 – CAIB Report (Columbia accident):**
  - Engineering solutions should have included a quantifiable range of uncertainty and risk analysis.
- **2005 – RTFTG Final Report (Columbia accident):**
  - Further compounding the modeling challenge is the fact that the models are deterministic, yielding point estimates, without incorporating any measure of uncertainty in the result
- **2005 – NASA CEV RFP**
  - Design and execute a meaningful risk mitigation program that culminates in a risk reduction flight effort and PDR by the end of calendar year 2008
- **2007 – NASA-STD-7009 for Models & Simulations:**
  - The risk assumed by the decision maker is often misestimated due to inadequate assessment of uncertainties
  - Reports to decision makers of M&S results shall include an estimate of their uncertainty and a description of any processes used to obtain this estimate

Uncertainty Analysis and Robust Design

Increase confidence and consistency in aerospace vehicle safety predictions by developing improved methods for <u>quantifying</u> and <u>managing</u> uncertainty

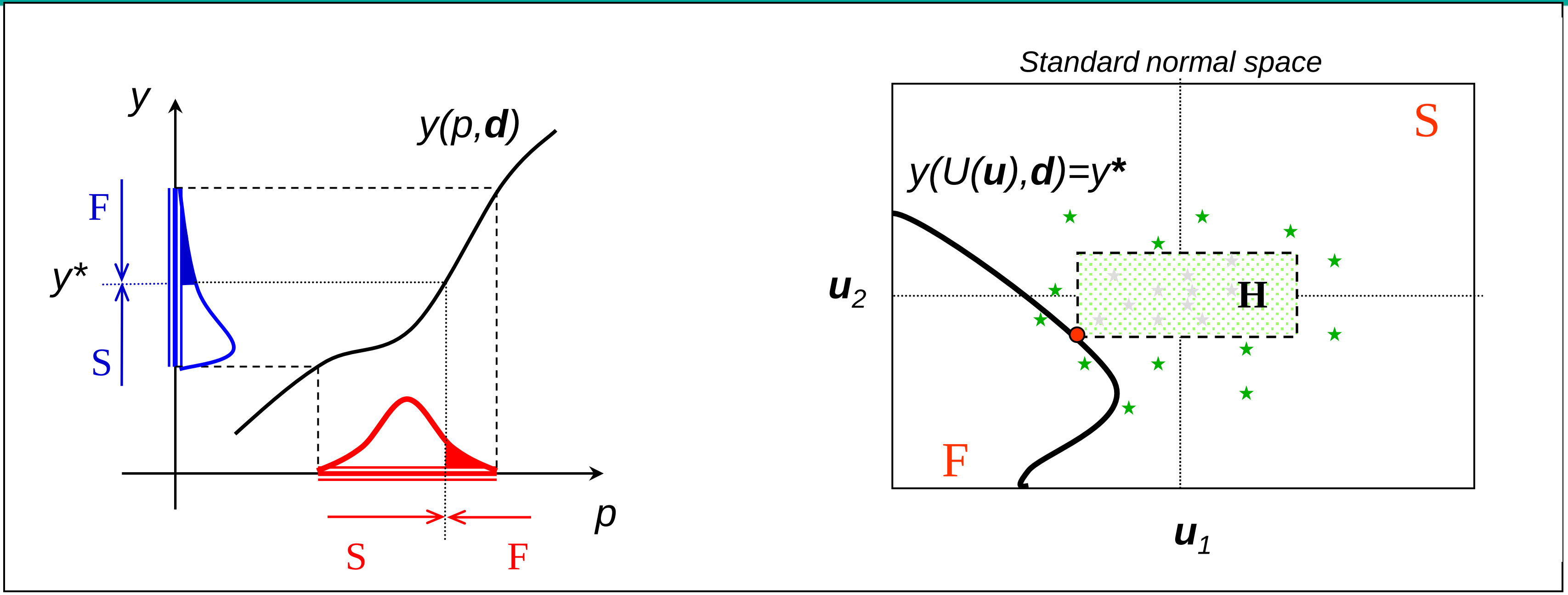
<u>Quantifying</u>

- Uncertainty Modeling
  - *Model uncertainty based on experimental data, simulations and/or expert opinion*
- Uncertainty Propagation
  - *Given uncertainty models of a system's inputs, how to propagate them through system models, to efficiently evaluate the corresponding system's outputs?*

<u>Managing</u>

- Robust Design
  - *Generate designs that robustly accommodate uncertainty*
- Uncertainty Decomposition
  - *Identify uncertainties that contribute the most to performance degradation*
  - *Determine the parameters that should (not) be modeled as uncertain*

Failure Domain Bounding via Homothetic Deformations

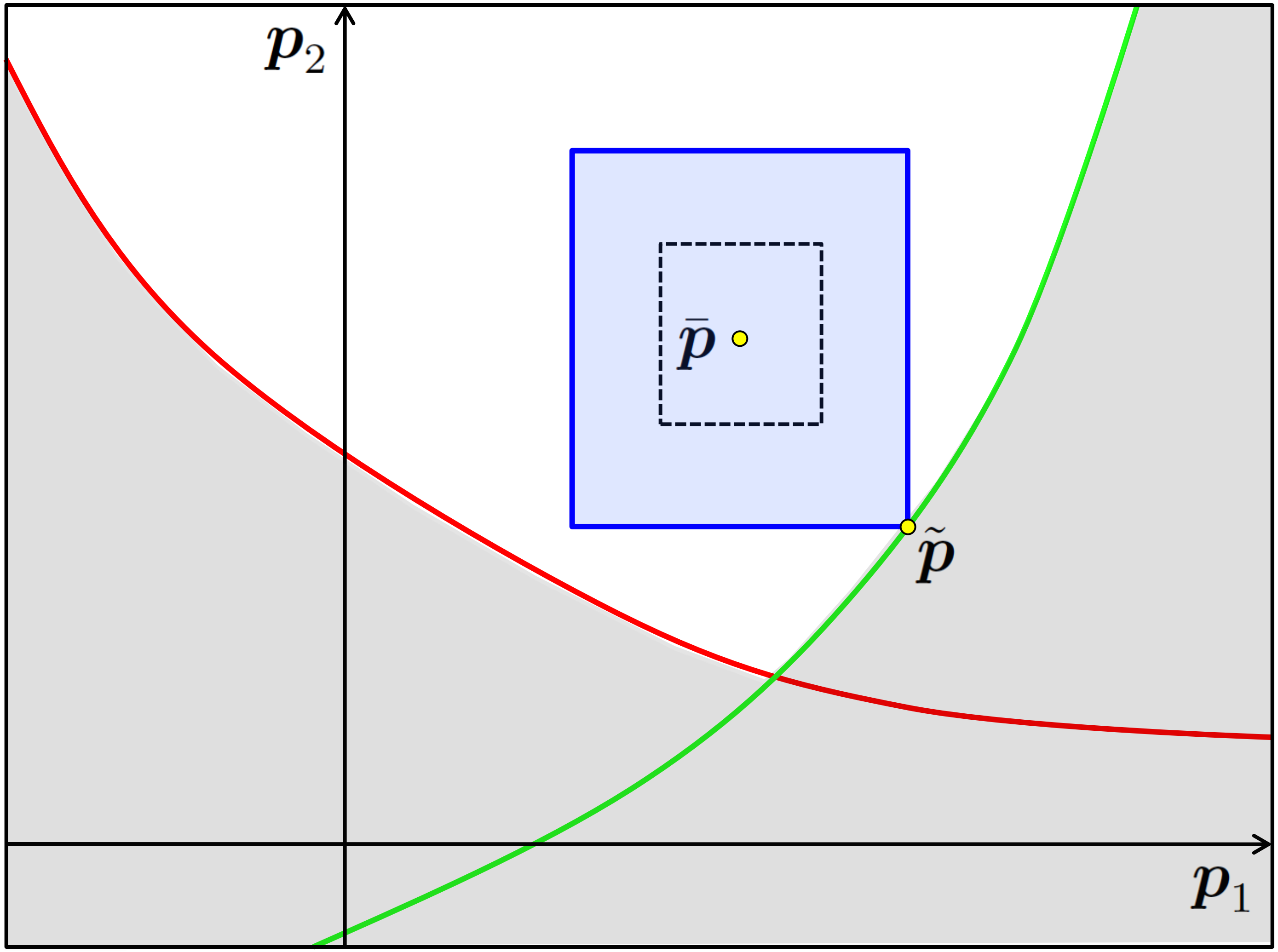


- Monte Carlo sampling
- Homothetic deformations
  - Optimization-based
  - Computationally cheap
  - Analytical expressions for P[H]

Applications

1. Robustness metric → PSM
2. Upper bounds to P[F]
3. Hybrid Method for P[F]

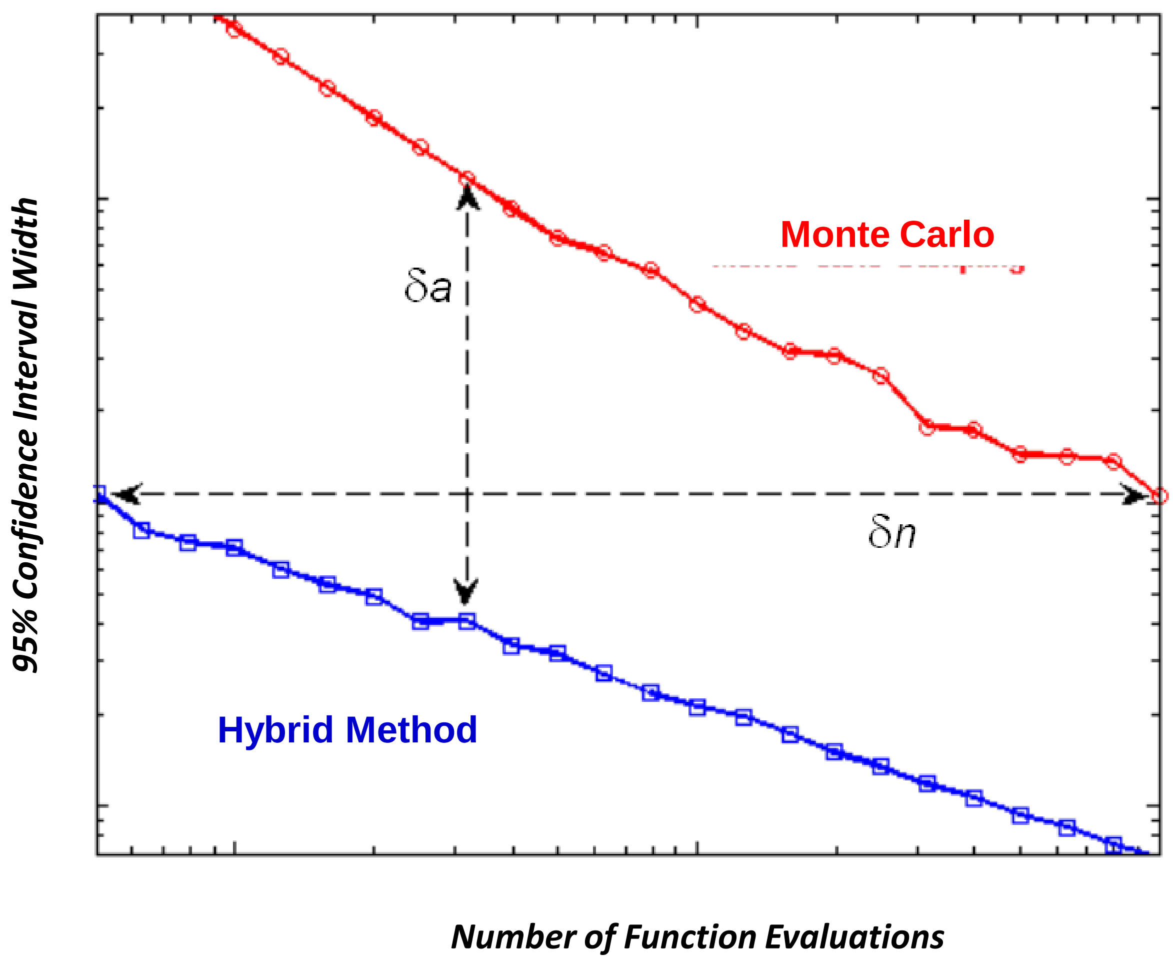
Analysis: Homothetic Deformations



- ***Outcomes: robustness metric, worst-case uncertainty, separation, probability bounds***

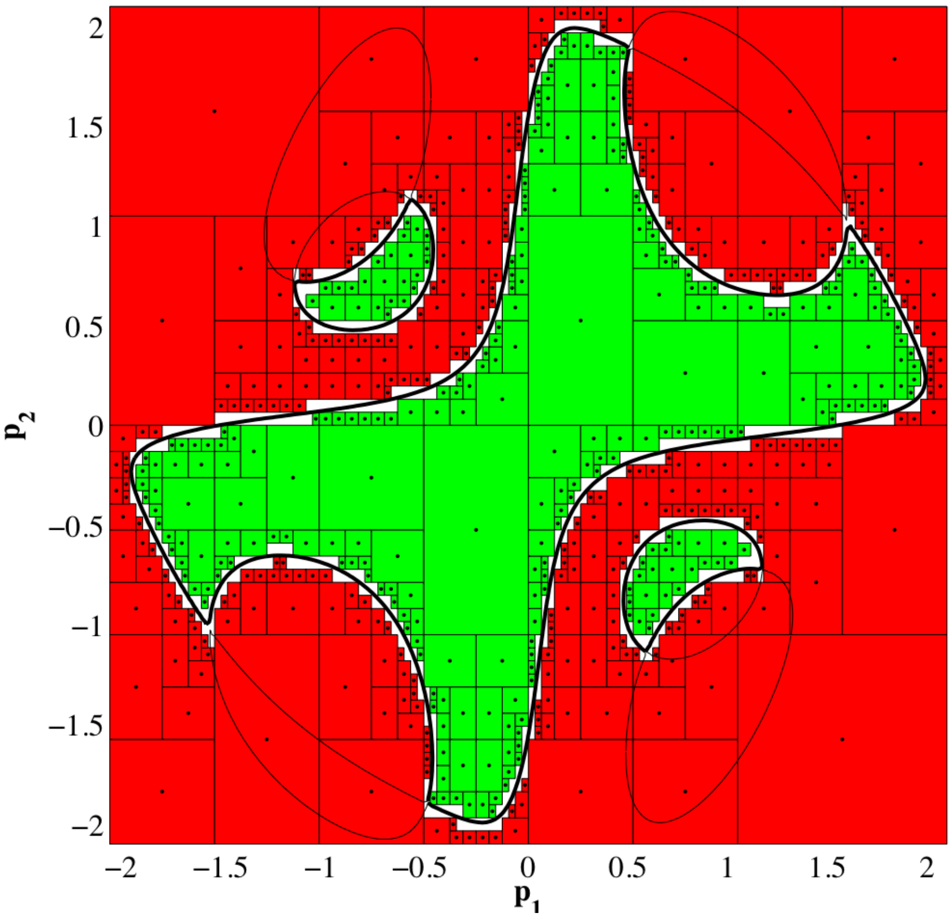
Efficiency Relative to Monte Carlo

Hybrid Method vs. Monte Carlo



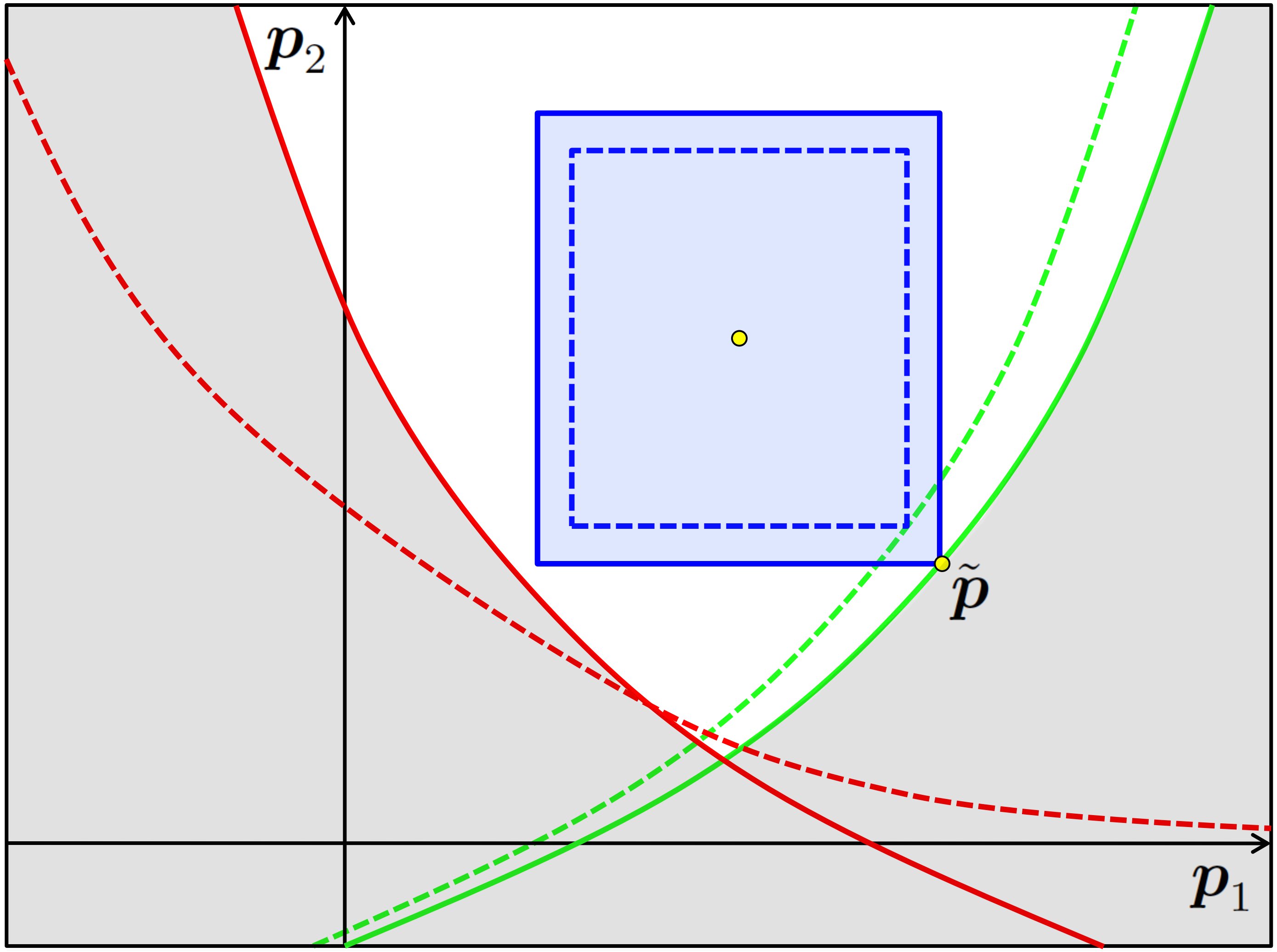
Failure Domain Approximation

- Builds upon the research in homothetic deformations.
- Yields high fidelity characterizations of complex nonlinear failure domains.
- Utilizes theory of Bernstein polynomials.
- Desensitizes the analysis from assumptions used to model the uncertainty.



Safe Domain ***Failure Domain***

Design: Homothetic Deformations



• ***Conflicting objectives, optimally robust designs***

The Tool Suite: UQTools

UQTools is a collection of Matlab functions designed to quantify the impact of uncertainty on generic, continuous, parameterized models.

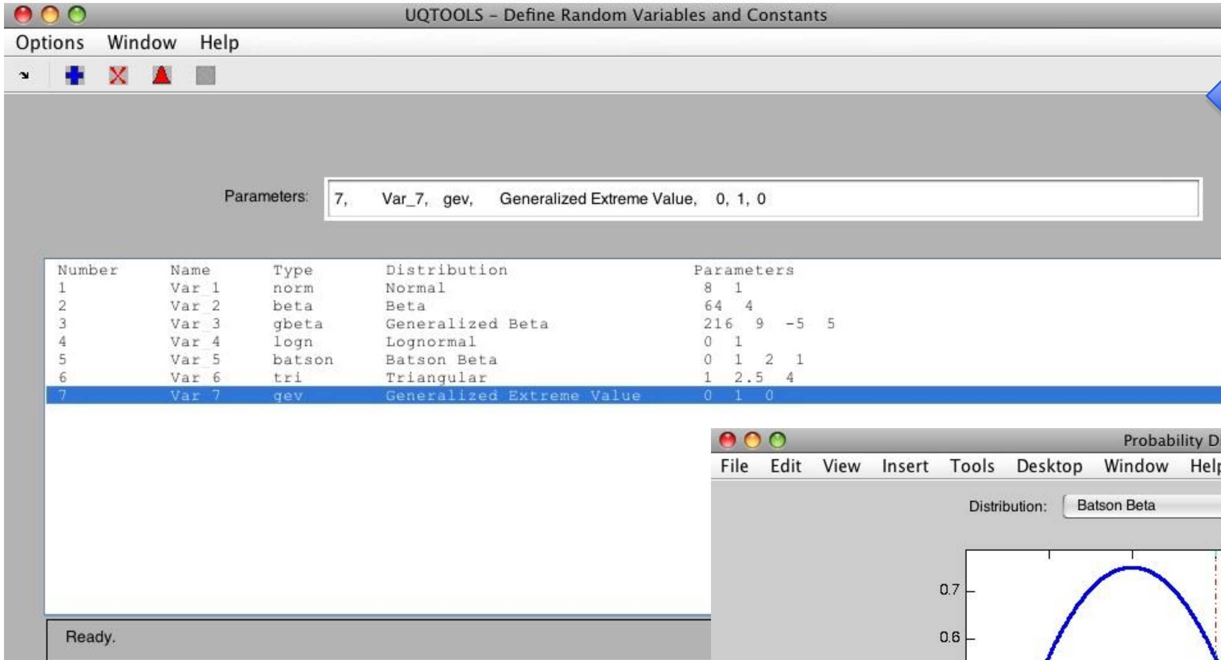
- Uncertainty Models:
  - Probability density functions
  - Non-probabilistic sets (hyper-rectangular and hyper-spherical)
- System Models:
  - Matlab-callable parameterized input/output maps.

UQTools Capabilities

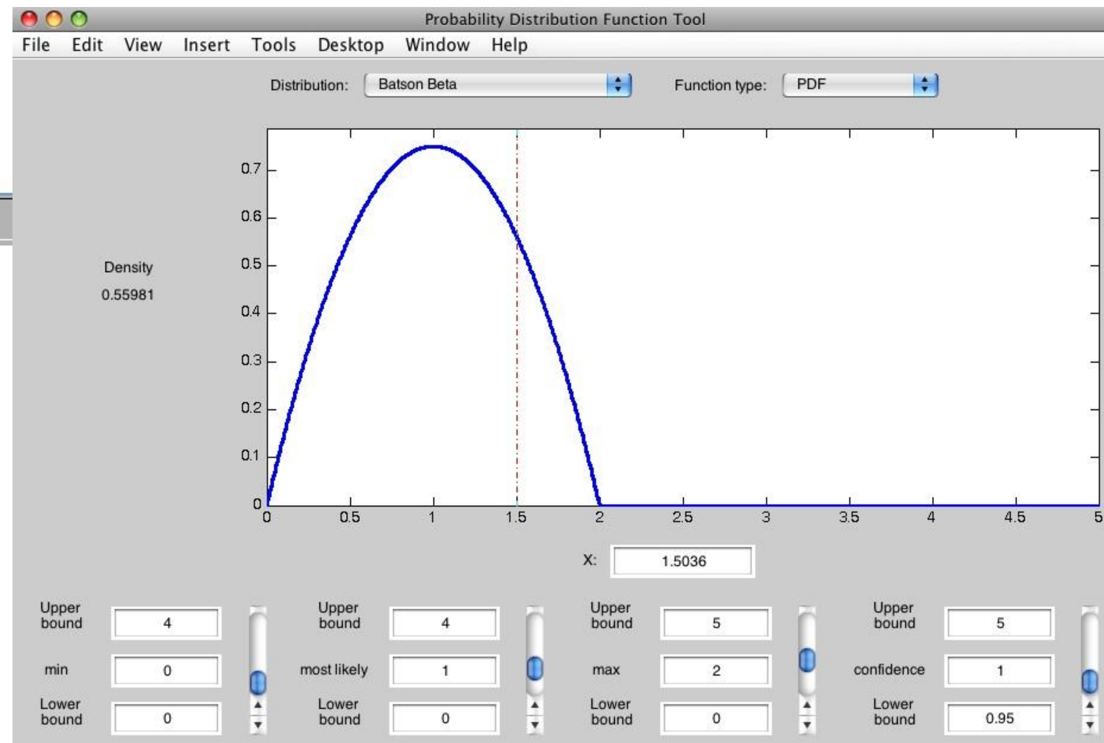
Integration of a collection of tools (each piece designed to attack a specific issue in UQ). The integration represents a unique capability in the field.

- Efficient methods for failure set bounding
  - Optimization-based approach for computing upper bounds on failure domains
- Hybrid methods for efficient estimation of failure probabilities
  - Combining failure set bounding theory with efficient conditional sampling
- First-Order Reliability Method
  - Efficient failure probability approximation for low probability 'tail' events
- Efficient deterministic sampling
  - Substantial improvement over conventional Monte Carlo
- Efficient moment propagation methods
  - Useful for propagating trends, e.g., mean & variance of system response
- Probabilistic sensitivity analysis
  - Analyze and rank the relative importance of system parameters
- Response surface tools
  - Radial basis functions and generalized polynomials (with 1^{st} and 2^{nd} derivatives)

Graphical User Interface



Master Parameter List Interface



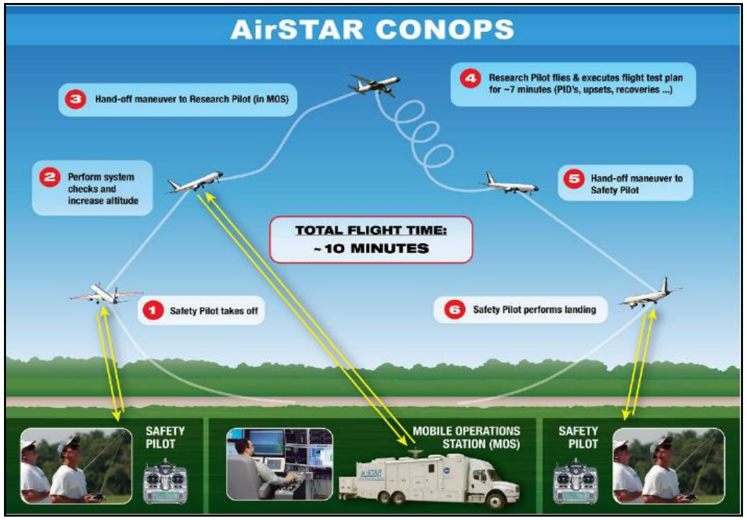
<u>Density Function Utility</u>

- ***Currently 20 different distributions supported.***
- ***General intervals and deterministic parameters also supported.***

Example Applications

Example: GTM Control Analysis & Design

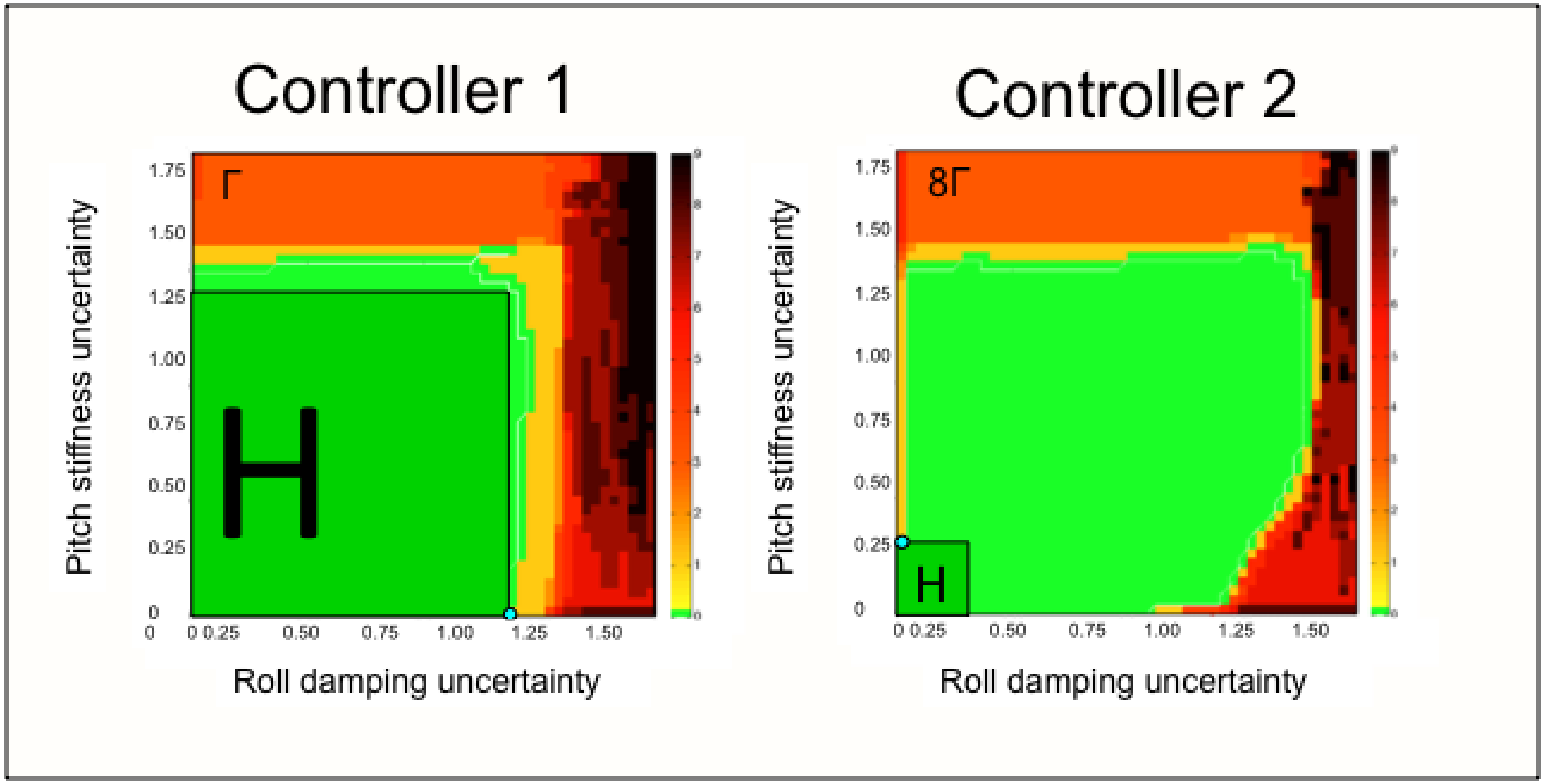
- Subscale, physical model
- High-fidelity Simulink model: non-linear aero, avionics, engine and sensor dynamics, atmospheric model, telemetry effects, time delay, filters, etc. (278 states)



- Control structure: LQR-PI and Model Reference Adaptive
- Uncertain parameters: aerodynamic coefficients
- Requirements: structural integrity, reliable flight envelope, command following, high frequency/residual oscillation

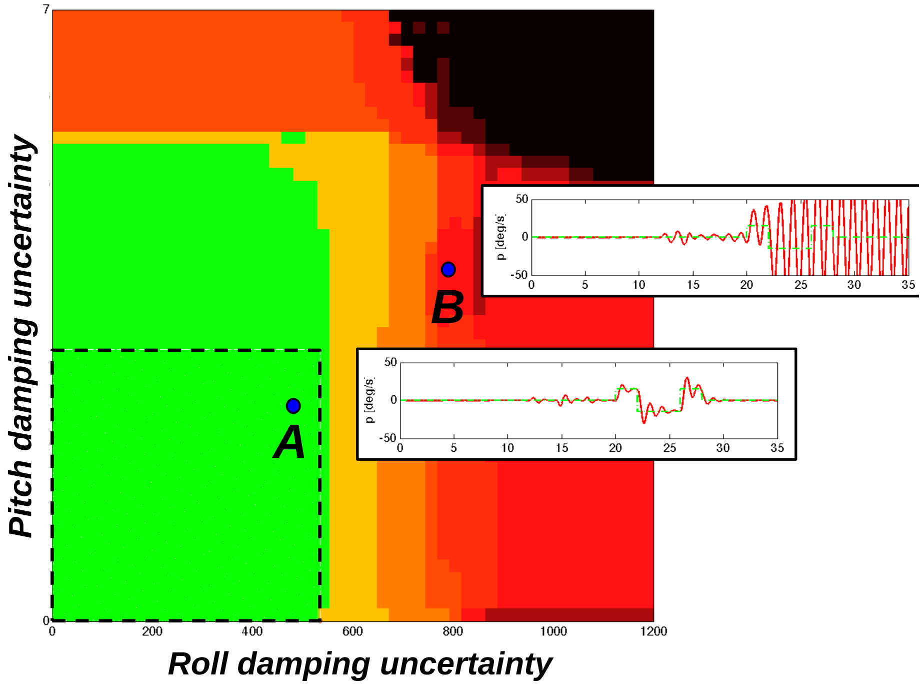
Homothetic Deformations: Analysis

- Two different controllers analyzed for robustness to aero uncertainties.
- Despite having a larger safe operating region, Controller 2 (high gain controller) has a nearly undetectable failure mode close to nominal point.
- Conventional Monte Carlo is not well-suited to capture this type of failure.

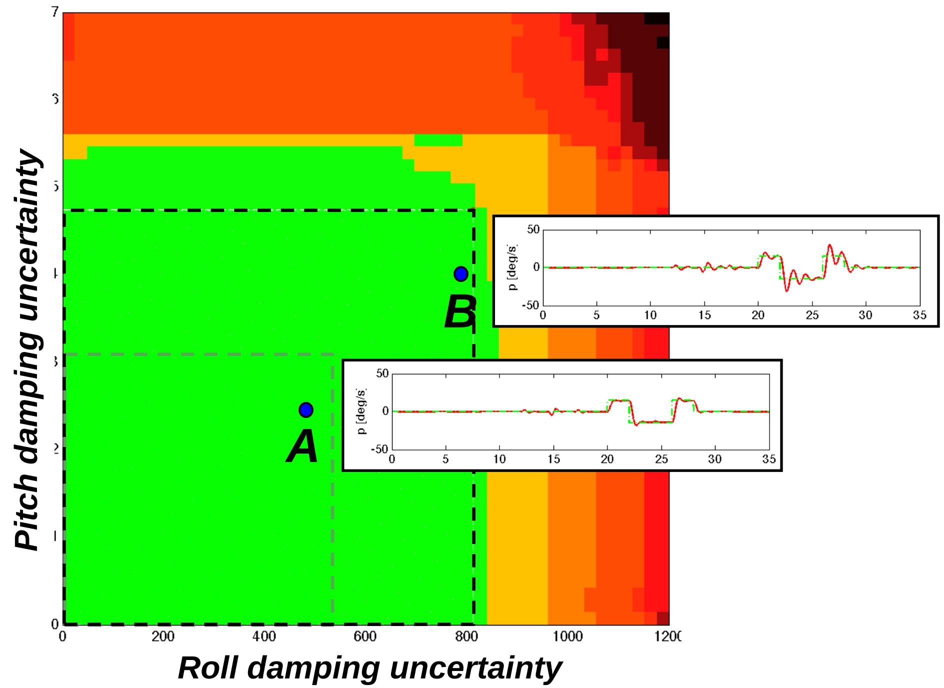


Homothetic Deformations: Design

Before Redesign

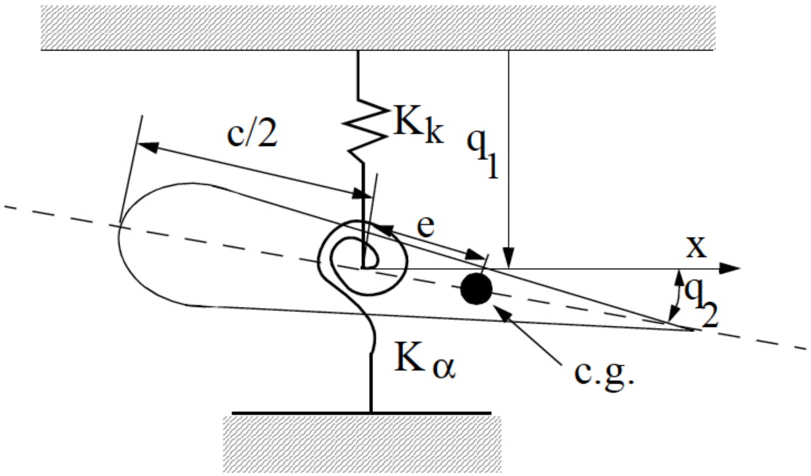


After Redesign



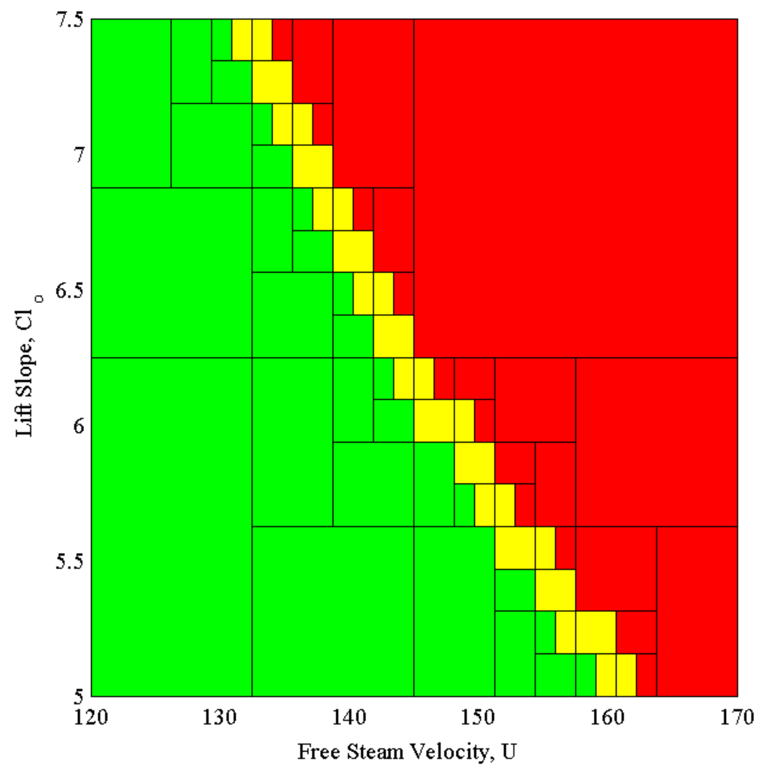
65 % improvement *in* roll *d*amping uncertainty

Simple Aeroelastic Model

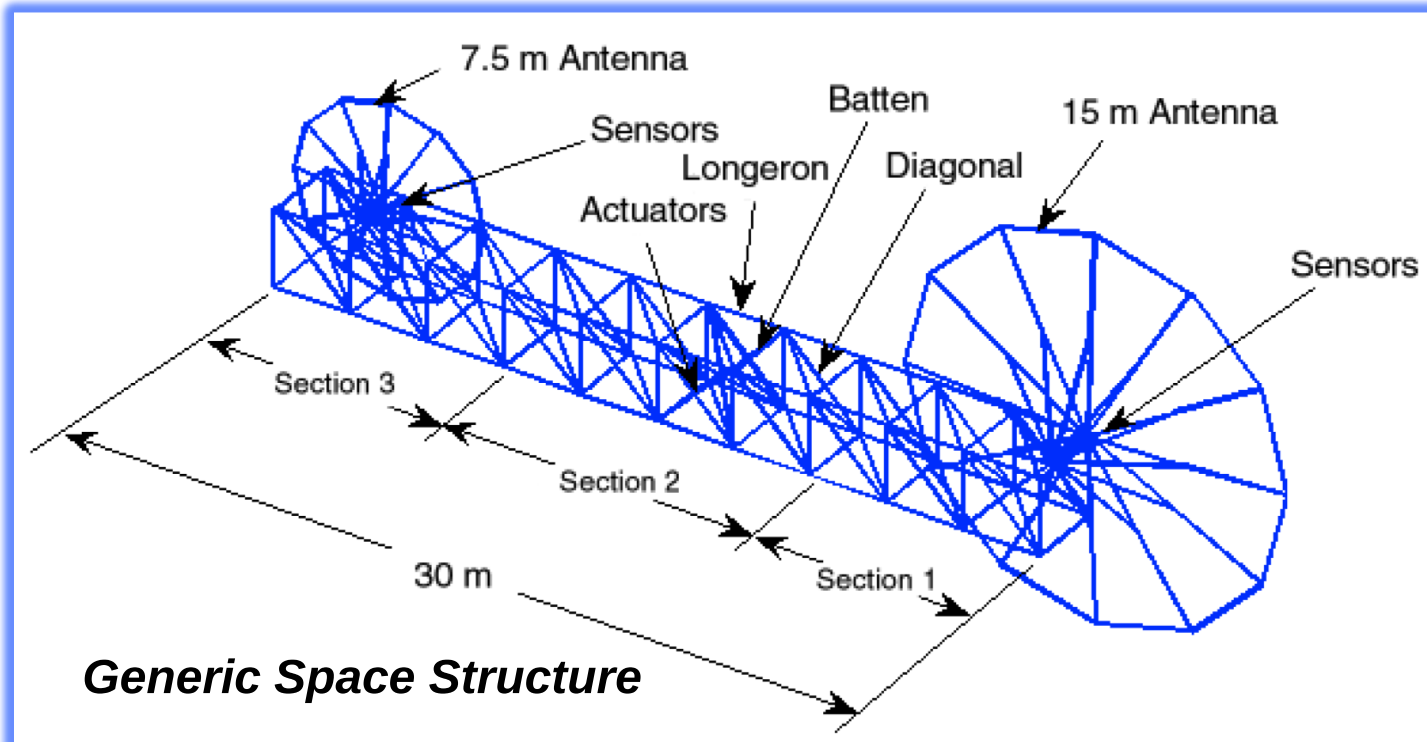


- Linear Aeroelasticity
- 4 uncertain parameters
- Bernstein polynomials
- 5 Routh-Hurwitz constraints

$$
A(U, K_k, K_\alpha, Cl_\alpha)\Big| = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \dfrac{-10K_k}{9} & \left(\dfrac{-245\,Cl_\alpha U^2}{288} + \dfrac{10K_\alpha}{9}\right) & \dfrac{-245\,Cl_\alpha U}{288} & 0 \\ \dfrac{10K_k}{9} & \left(\dfrac{343\,Cl_\alpha U^2}{144} - \dfrac{100K_\alpha}{9}\right) & \dfrac{343\,Cl_\alpha U}{144} & 0 \end{bmatrix}
$$

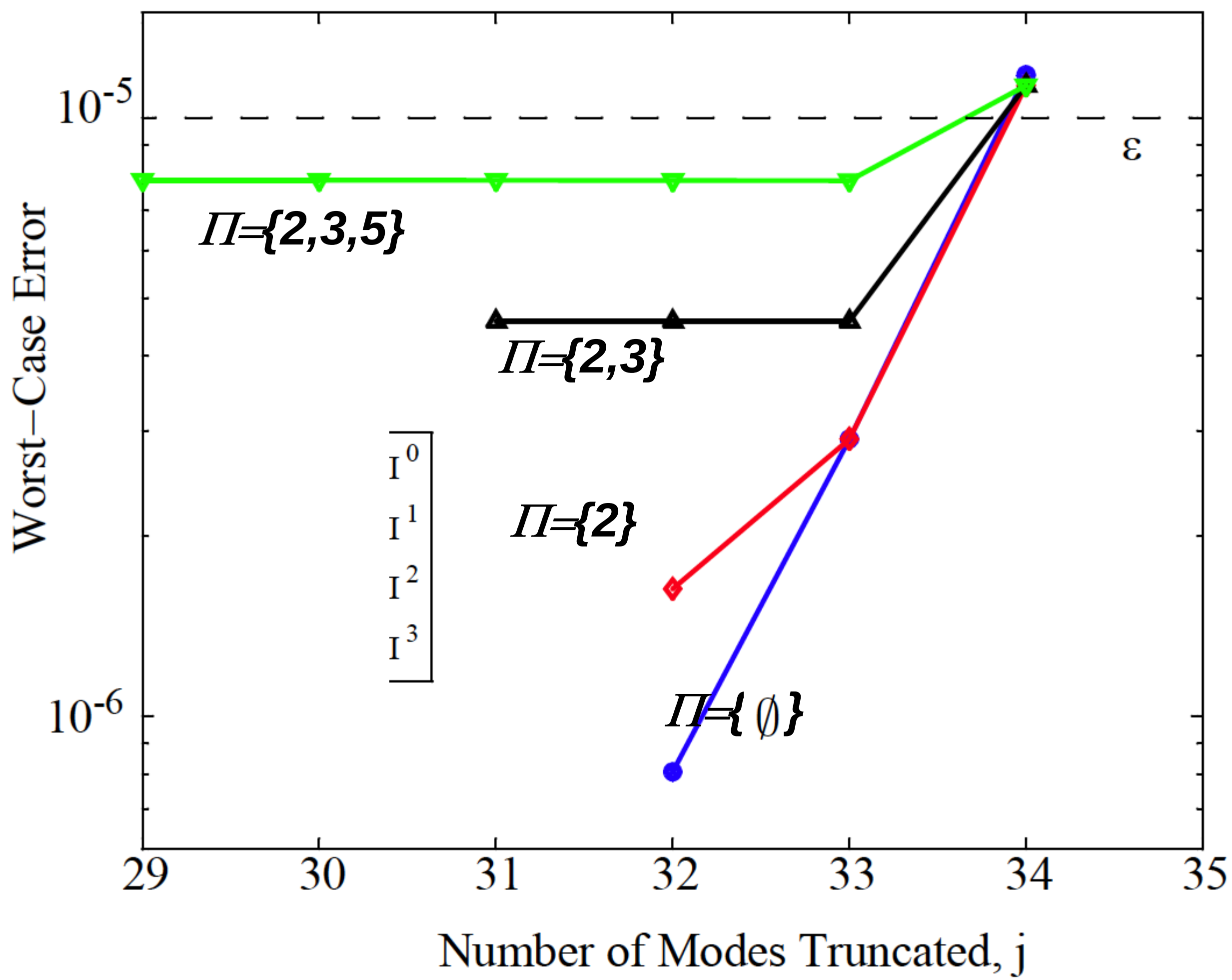


Example: Flexible Antenna



- Uncertain parameters: outer radius of precision tube members (14 groups, with Δp = +/- 0.5 %)
- Maximum allowable error, $\varepsilon=10^{-5}$, (-100 dB)
- MIMO system (78 states, 3 inputs, 6 outputs)

Example: Flexible Antenna



Efforts to Improve State of Practice

- In June 2009, LaRC formed an Uncertainty Based Methods Community Forum (currently 60 members – one AMES CS + one private sector researcher)

  The objective of the UBM Community Forum is to facilitate the cross fertilization of methods, tools, and ideas related to statistical and probabilistic analysis and design for a broad spectrum of engineering applications

- Foster informal and open discussion of problems in uncertainty based methods
- Discuss success and/or failure
- Share current applications
- Provide a forum for new methods to be vetted

Conclusions

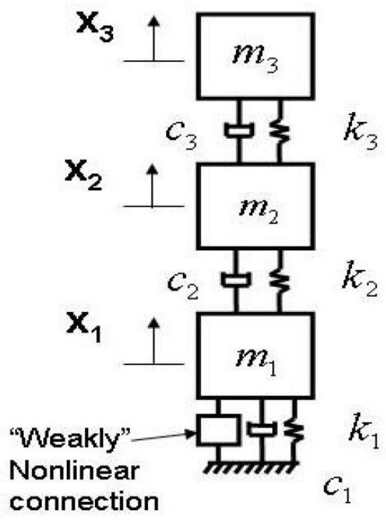
- ***High-fidelity characterization of the failure domain***
- ***Identification of worst-case uncertainty combinations***
- ***Exact failure probability bounds***
- ***Substantially desensitizes the uncertainty analysis from the uncertainty model assumed***

Validation Challenge Workshop

Sandia National Laboratory Workshop

(13 international teams of experts chosen to participate)

Challenge: Adequate statistical characterization of uncertainty using limited experimental data.

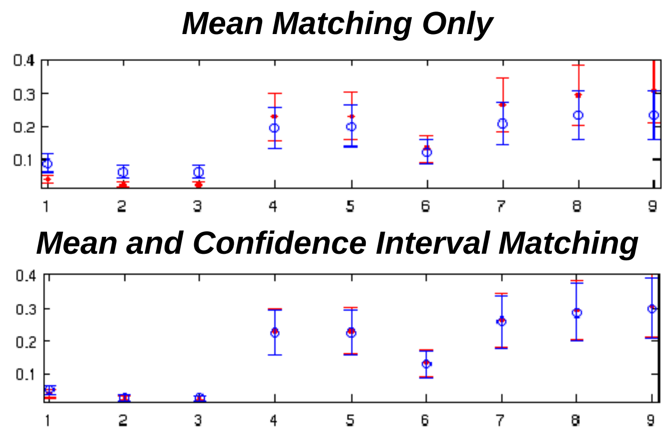


Uncertain/Nonlinear Subsystem

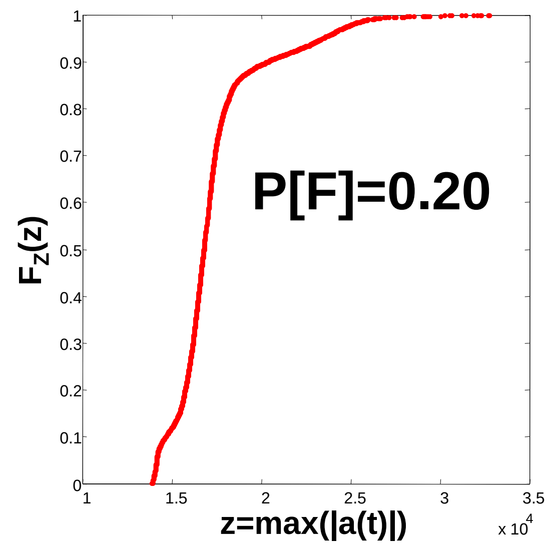
40 data points (2 sets of 20)

LaRC's Solution

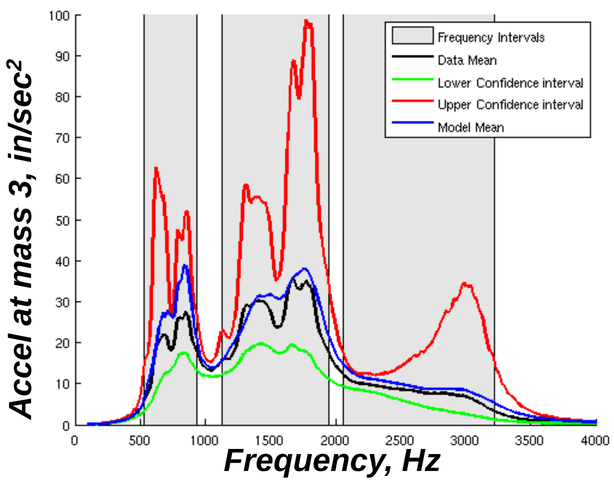
Statistical matching of parametric uncertainties



Results on Target System



Statistical response of uncertainty model versus data



Regulatory Requirement

P[z>18000 in/sec^2] < 0.01

An Empirical Study of Variables Acceptance Sampling: Methods, Implementation, Testing, and Recommendations

K. Preston White, Jr.

Professor of Systems and Information Engineering, University of Virginia

Visiting Professor of Operations Research, Naval Postgraduate School

Member, NASA Engineering Statistics Team

Kenneth L. Johnson

Statistics and Trending

NASA/ NESC Systems Engineering Office

NASA Statistical Engineering Symposium

04 May 2011

Overview

- Acceptance sampling/Sampling plans
- Motivation
- Components of a probabilistic requirement
- Current NASA best practice (ASA)
- A potentially more efficient practice (ASV)
- Research plan, summary results, literature review
- Operating characteristic
- Derivation of variables sampling plans
- ASV sampling plan calculators
- Empirical testing and results
- Tests of the fundamental assumption (near normal, near exponential skew, and modest skew)
- Procedure for selecting a sampling plan (flow diagram
- Summary/Contributions

Acceptance sampling

- One of the oldest problems in quality engineering is to assess the acceptability of items that a customer receives from a producer.
- Acceptance sampling is an alternative to 100% inspection applied when inspection is destructive, or when the time and/or cost of 100% inspection are unwarranted or prohibitive.
- Based on inspection of the sample, the customer decides whether to accept or reject the entire lot, or to continue sampling.
- There are standards (MIL, ANSI, and ISO) pertaining to acceptance sampling.

Sampling plans

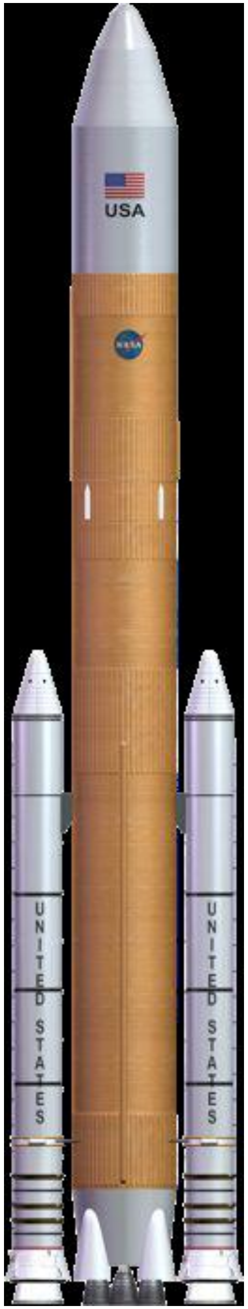
- A sampling plan is the pair (n,c) or (n,k) , where n is the minimum sample size, i.e., the minimum number of observations required to verify statistically the requirement.
- For discrete random variables, the constant c is the maximum number of nonconforming observations supporting the determination that a lot is acceptable.
- For continuous random variables, constant multiplier k is the minimum distance (in standard deviations) between the sample mean and the required limit supporting the determination that a lot is acceptable.

Motivation

Our interest in acceptance sampling arose in an analogous sampling experiment--the need to verify level-two design requirements for Cx “by analysis” using Monte Carlo simulation.

[CA0049-PO] The CaLV [Cargo Launch Vehicle] shall launch LSAM [Lunar Surface Access Module] from the launch site to the Earth Rendezvous Orbit (ERO) for Lunar Sortie Crew and Lunar Outpost Crew missions

The delivery of the LSAM from the launch site to the ERO shall be verified by analysis. The analysis shall be performed using NASA-accredited digital flight simulations. The analysis shall include Monte Carlo dispersions on mass properties, engine performance, GN&C parameters and environmental parameters. The verification shall be considered successful when the analysis results show that there is a [ρ] probability with a [100(1-β)%] confidence that the LSAM reaches ERO.



Components of a probabilistic design requirement

- Condition (I)

  conformance indicator (typically a limit on the value of an output variable)

- Reliability (ρ)

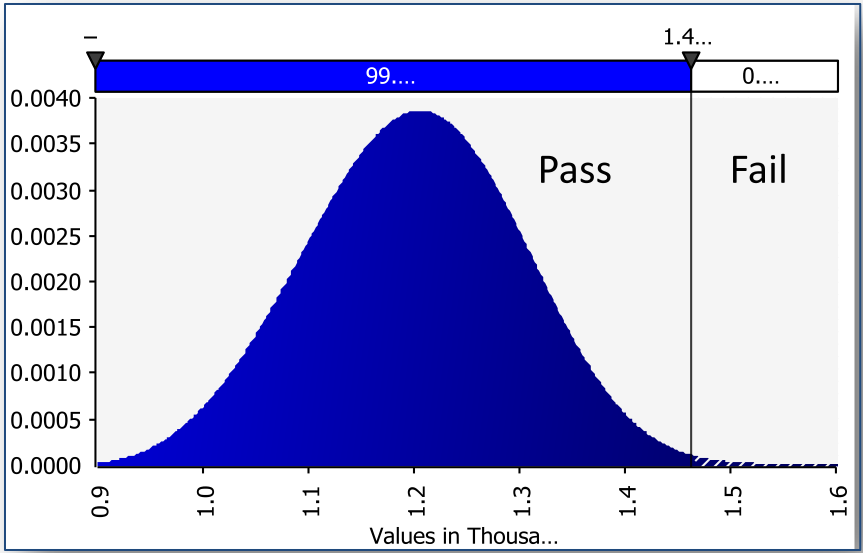
  minimum probability of achieving the condition

- Consumer's risk (β)

  maximum probability of accepting a nonconforming design

- Producer's risk (α)

  maximum probability of rejecting a conforming design

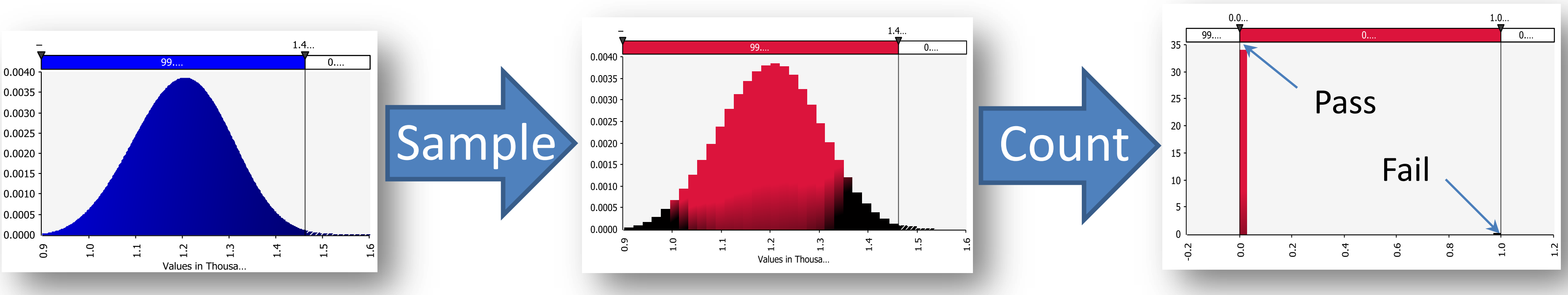


Consider the (true but unknown) parent distribution of an output variable X.

We can see that this output meets the condition $X \leq 1463$ with reliability ρ=0.997.

If we knew the parent distribution a priori, there would be no sampling error and the risks would be $\beta=\alpha=0$.

Current NASA best practice for requirements verification using Monte Carlo



Sample the parent distribution using Monte Carlo simulation

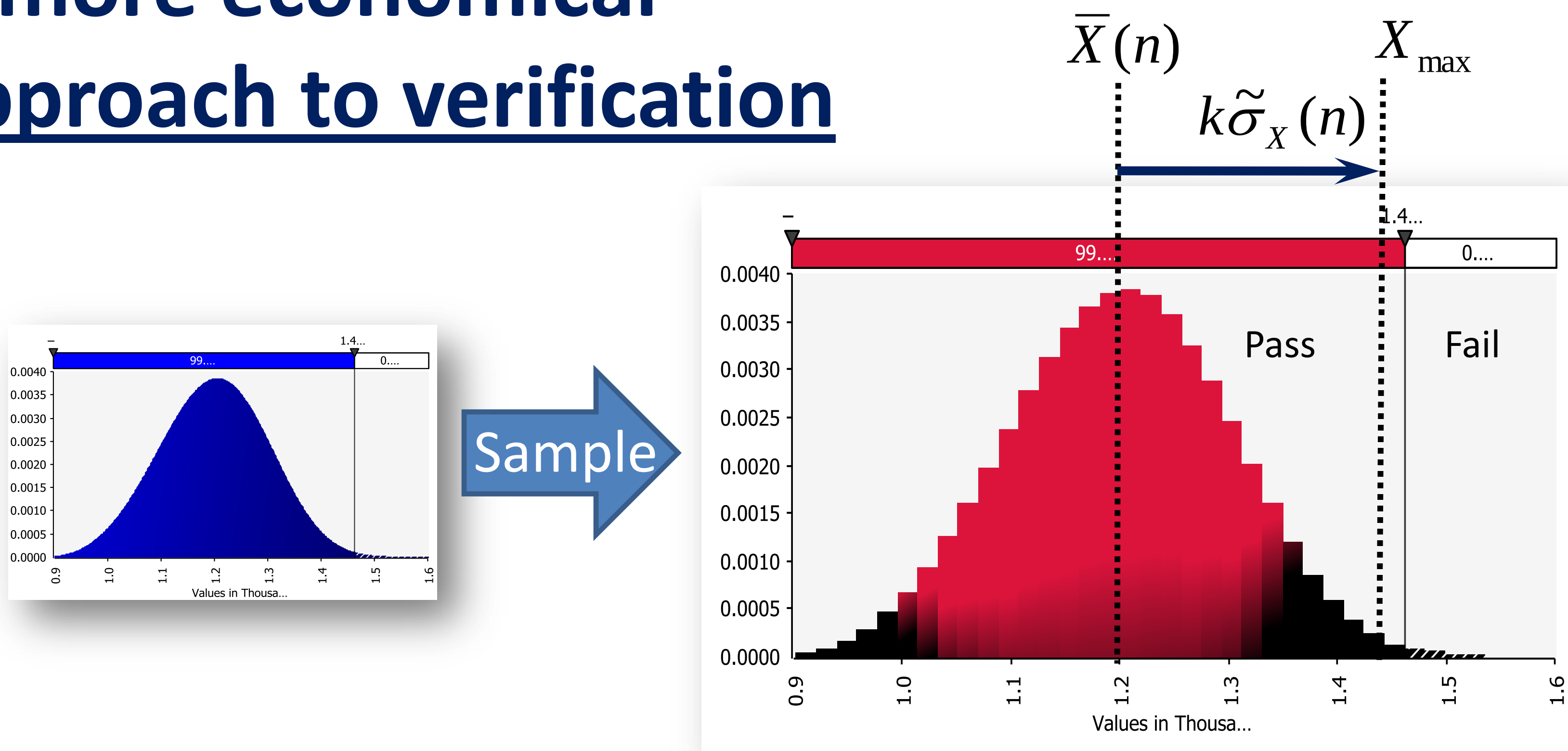
Count the number of simulation trials in which the output fails the condition

The current best practice employs **attributes acceptance sampling** (ASA). For the required reliability and consumer's risk, the sampling plan specifies **number of trials** (n) and the **maximum number of failures permitted** (c) to substantiate the validity of the design.

<u>Advantage:</u> Plans are exact and can be determined a priori. (Nonparametric--by definition, the distribution of the count is Binomial(n,p), where p is the true reliability.)

<u>Disadvantage:</u> Plans require large samples for high confidence in highly reliable designs (the pass/fail count ignores "by how much").

A more economical approach to verification



The current project seeks a best practice employing **variables acceptance sampling** (ASV). For the required reliability and consumer's risk, the sampling plan specifies **number of trials** (n) and the **minimum multiplier** (k) to substantiate the validity of the design.

<u>Advantage:</u> ASV plans typically require fewer trials than ASA plans (but not always).

<u>Disadvantages:</u> Software for plan generation is unavailable; procedures/assumptions reported in the academic literature appear to be largely untested.

Research plan and summary results

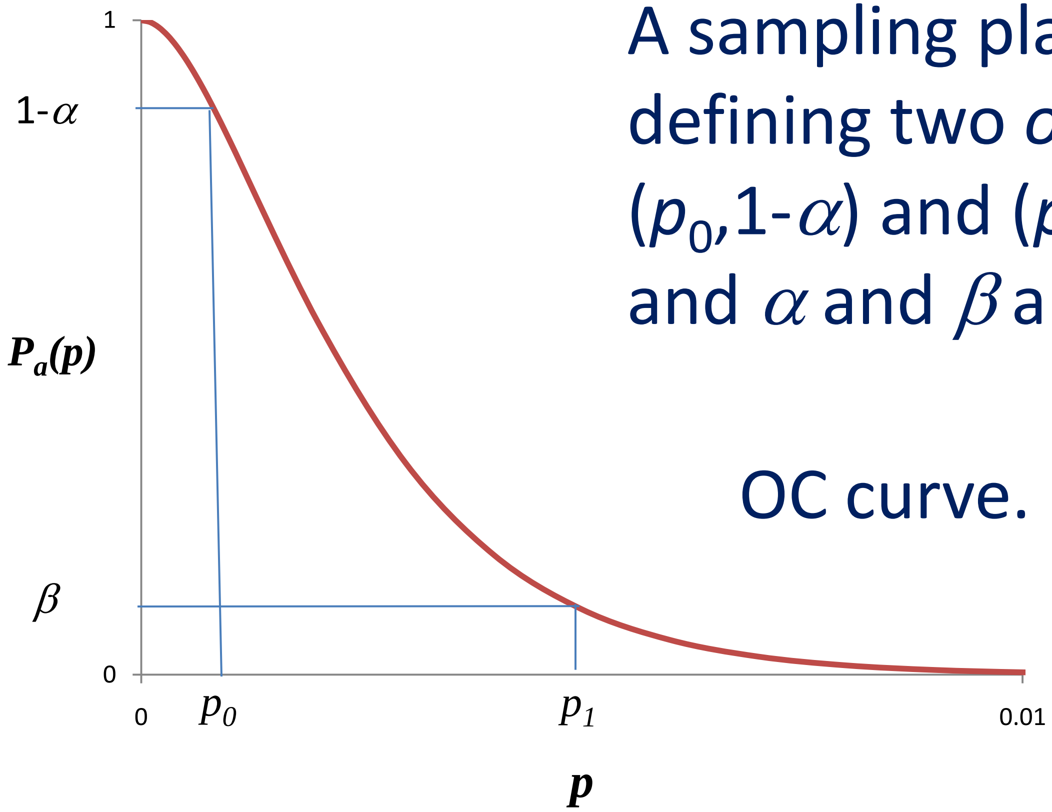
Software search	Off-the-shelf plan calculators (commercial or otherwise) were found only for normal variates.
Literature search	Plans for 5 additional variates were found in the academic literature (Exponential, Gamma, Weibull, Inverse Gaussian, Poisson, Burr).
Implementation	Calculators were implemented in Excel for Binomial, Normal, Exponential, Gamma, Weibull, Inverse Gaussian, and Poisson. (Burr not attempted.) Verified against published examples. Plans typically, but not invariably, smaller than corresponding ASA plans.
Empirical Testing	Monte Carlo simulation applied to test plans derived for typical (Constellation-like) OC from all seven calculators. All were validated, except for Inverse Gaussian. Error in the published IG derivation discovered.
Application issues	Fundamental assumption that the distributional form can be determined uniquely from sample data tested using Monte Carlo. Assumption not substantiated for typical OC. Conservative protocol developed for selecting plans to use in practice.

Literature

Variable	Source	Implemented	Validated
Binomial	Multiple sources	✓	✓
Normal	Multiple sources	✓	✓
Gamma	K. Takagi (1972) "On designing unknown-sigma sampling plans on a wide class of non-normal distributions," *Technometrics* 14(3)669-678.	✓	✓
Weibull	K. Takagi (1972) "On designing unknown-sigma sampling plans on a wide class of non-normal distributions," *Technometrics* 14(3)669-678.	✓	✓
Exponential	W. C. Guenther (1977), *Sampling Inspection in Statistical Quality Control,* Macmillan, New York.	✓	✓
Poisson	W. C. Guenther (1977), *Sampling Inspection in Statistical Quality Control,* Macmillan, New York.	✓	✓
Inverse Gaussian	M. S. Aminzadeh (1996), "Inverse-Gaussian Acceptance Sampling Plans by Variables," *Communications is Statistics--Theory and Methods 25(5)923-935.*	✓	×
Burr	K. Takagi (1972) "On designing unknown-sigma sampling plans on a wide class of non-normal distributions," *Technometrics* 14(3)669-678.	No	No

Operating Characteristic

Every sampling plan has an *operating characteristic* (OC) which defines the probability of accepting a population $P_a(p)$ for every value of the failure probability $p \in [0,1]$.



A sampling plan is derived by defining two *operating points*, $(p_0, 1\text{-}\alpha)$ and (p_1, β), where $p_0 < p_1$ and α and β are small probabilities.

OC curve.

Derivation of variables plans

- The underlying problem can be framed as an hypothesis test for which we intend to enforce both significance and power requirements.
- The null and alternate hypotheses are

$$\boldsymbol{H_0}: p = p_0 \text{ and } \boldsymbol{H_1}: p = p_1 > p_0$$

- Under $\boldsymbol{H_0}$ we accept the population as conforming and under $\boldsymbol{H_1}$ we reject the population as nonconforming.
- The inequalities

$$P_a(p_0) \geq 1\text{-}\alpha \text{ and } P_a(p_1) \leq b$$

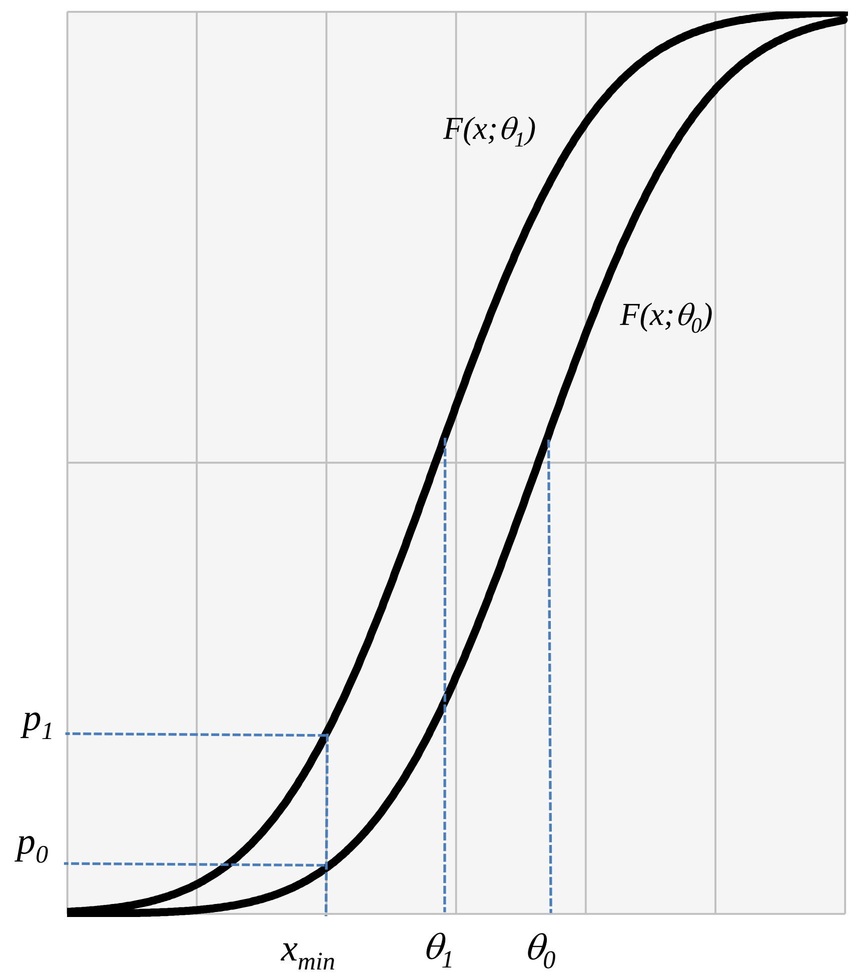
establish the significance and power of the test.

Derivation of variables plans

With the form of the distribution $F(x;\theta)$ known, the null and alternate hypotheses are equivalent to

$$H_0\!: \theta = \theta_0$$
$$H_1\!: \theta = \theta_1 > \theta_0$$

as shown for a required lower bound $x_{\min}$.

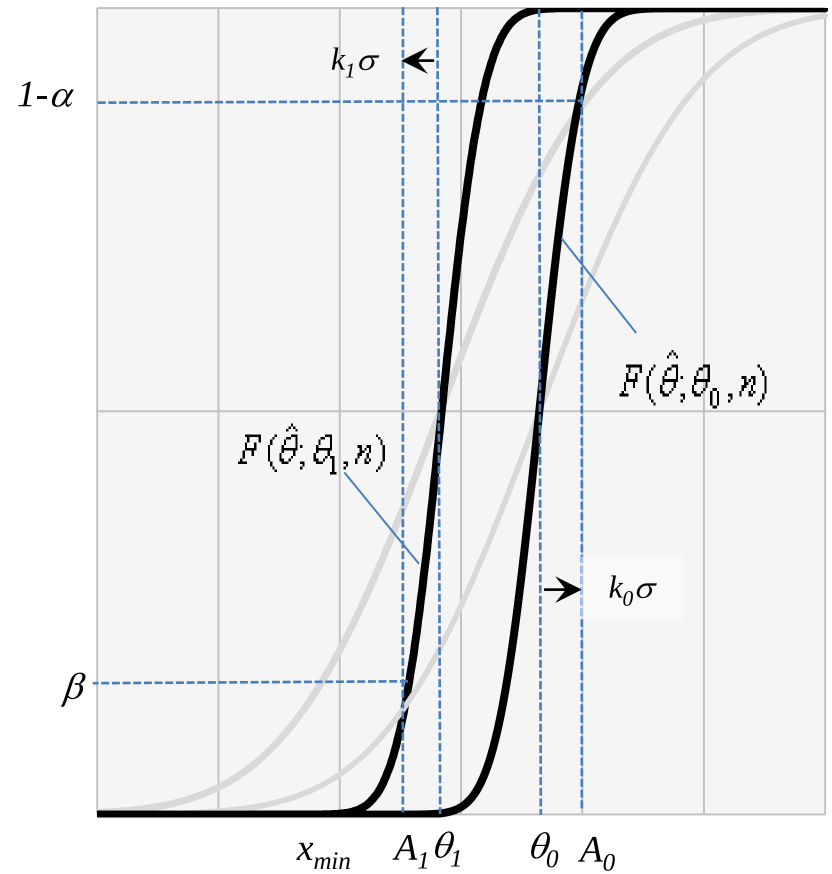


Derivation of variables plans

The power requirements are applied to the sampling distribution

$$F(\hat{\theta};\theta,n)$$

to determine the acceptance limit A, required sample size n, and multiplier k.



Dashboard for the Weibull Calculator

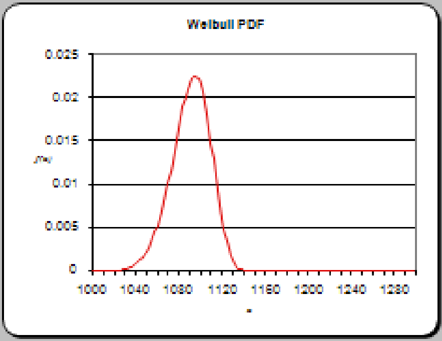
Weibull(ν,λ,γ) Acceptance Sampling Plans

rev KPW 18 April 2009

ENTER the parameters of the Weibull(ν,λ,γ) distribution in the fields below. The statistics for the Weibull random variable X are computed by the calculator.

Input	ν (shape)	λ (scale)	γ (location)
from fit	6	100	1000

Calculated	Mean μ	Std. Dev. σ	Skew	Kurtosis
from fit	1092.7719	17.976749	-0.373262	3.0354553



ENTER the lower and upper test points (reliabilities $0 \le p_l < p_u \le 0.999$), fixed β risk, and the lower or upper limit on the Weibull random variable X based on the requirement.

Input	p_l	β	p_u	X_{min} or X_{max}
requirement	0.9973	0.1	0.999	1136.7

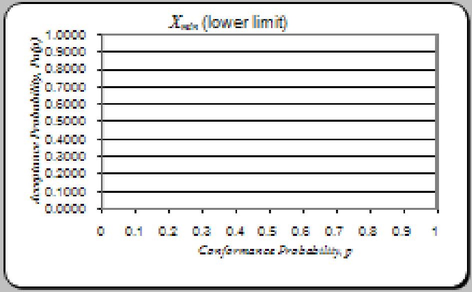
If the limit is less than the mean, the value entered is ssumed to be a lower limit; if the limit is greater than the mean, the value entered is assumed to be an upper limit.

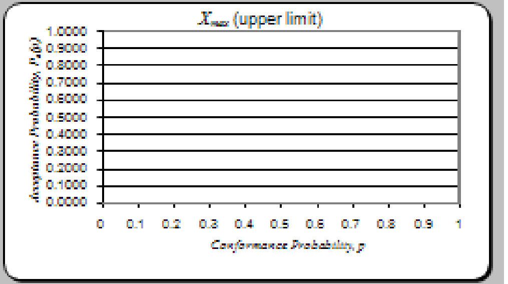
The corresponding (n, k) sampling plans and associated α risks are tabluated .

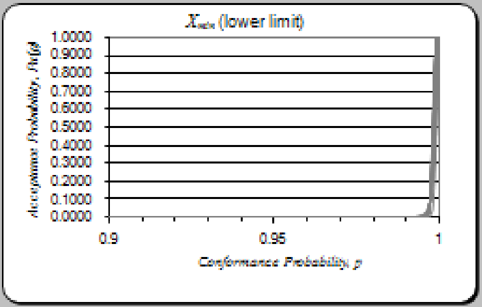
Results	n	k	α	Accept	$\mu_{critical}$
	1486	2.3775087	0.001	Yes	1135.5118
	1348	2.380439	0.002	Yes	1135.5645
	1267	2.3823936	0.003	Yes	1135.5996
	1209	2.383911	0.004	Yes	1135.6269
	1163	2.3851735	0.005	Yes	1135.6496
	1126	2.3862671	0.006	Yes	1135.6693
	1095	2.3872393	0.007	Yes	1135.6867
	1067	2.3881198	0.008	Yes	1135.7026
	1043	2.3889281	0.009	Yes	1135.7171
	1021	2.389678	0.01	Yes	1135.7306
	876	2.3953703	0.02	Yes	1135.8329
	790	2.3994907	0.03	Yes	1135.907
	728	2.4029002	0.04	Yes	1135.9683
	679	2.4058971	0.05	Yes	1136.0221
	639	2.4086241	0.06	Yes	1136.0712
	605	2.4111618	0.07	Yes	1136.1168
	576	2.4135606	0.08	Yes	1136.1599
	549	2.4158546	0.09	Yes	1136.2011
	526	2.4180679	0.1	Yes	1136.2409
	433	2.4284349	0.15	Yes	1136.4273
	366	2.4383814	0.2	Yes	1136.6061
	313	2.4484937	0.25	No	1136.7879
	269	2.4591699	0.3	No	1136.9798
	231	2.4707793	0.35	No	1137.1885
	197	2.4837409	0.4	No	1137.4215
	168	2.4985953	0.45	No	1137.6886
	141	2.5161042	0.5	No	1138.0033

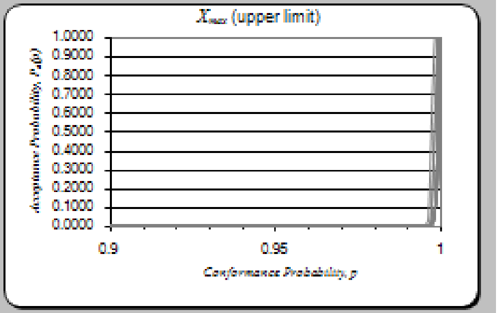
DETERMINE if the design is acceptable from the plan with the largest value of n which is no greater than the number of data points used to fit the distribution.

Operating Characteristic (OC) or Power Curves for all plans.





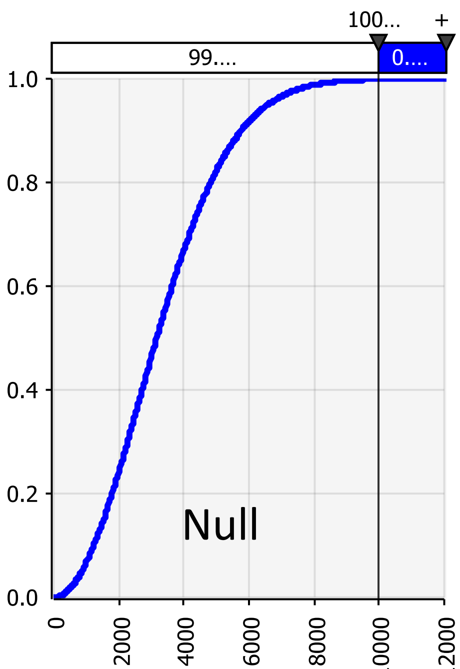


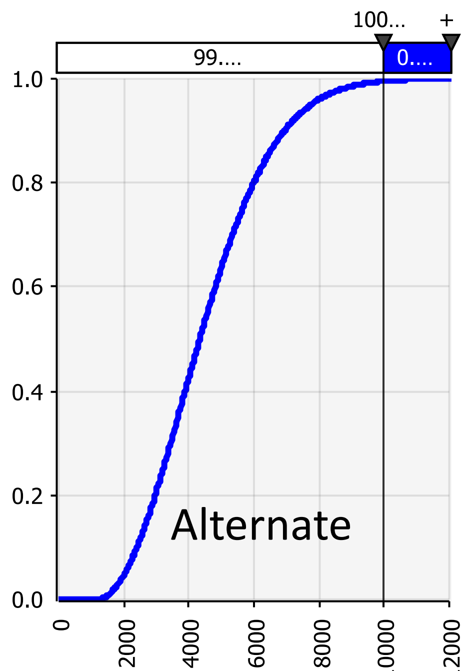


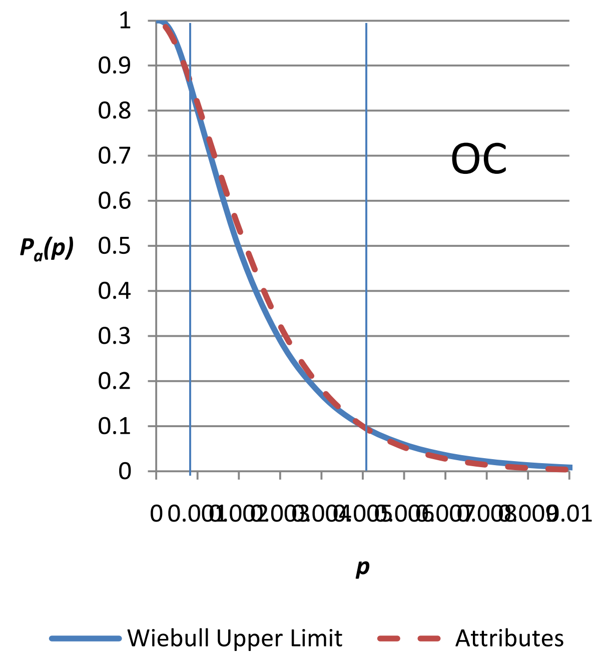
Calculations are aproximate, based on the procedure given by K. Takagi (1972) "On deigning uknown-sigma sampling plans based on a wide class of non-normal distributions, *Technometrics*, 14(3):669-678.

Empirical test results (example)

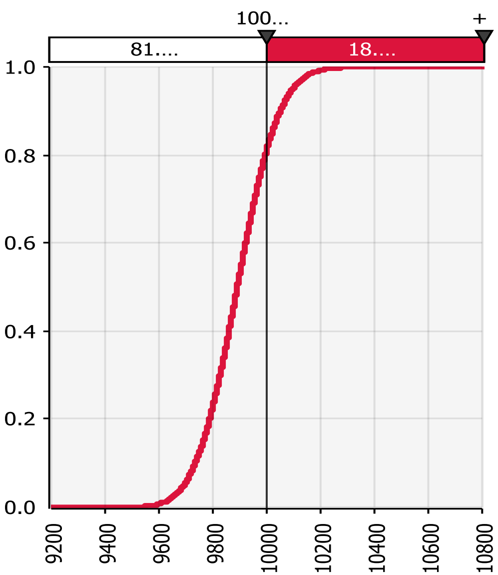
Consider an upper limit of x_{max}=10,000 for a random variable X distributed Weibull with unknown shift parameter δ and estimated shape and scale parameters. For the test OC (p_0,α)=(0.005,0.2), (p_1,β)=(0.001,0.1), the associated the null and alternative means are μ_0 =7841.64 and μ_1 =8121.07, respectively. The variables plan from the gamma calculator is (n,k)=(156, 2.17779).

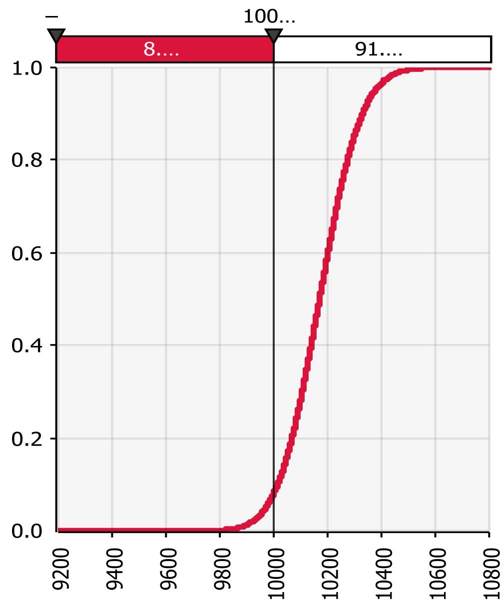




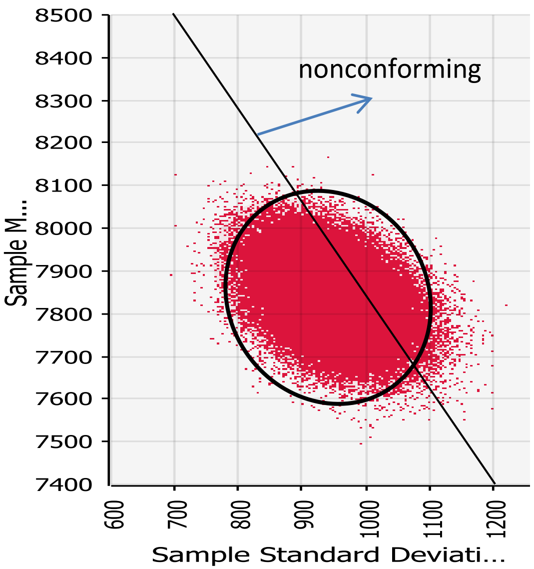


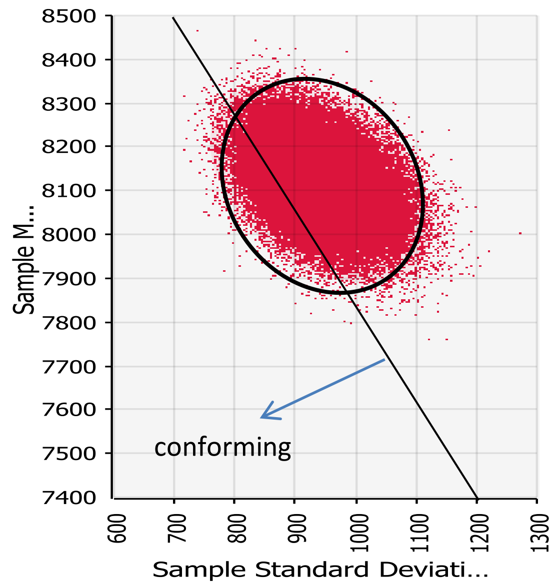
Empirical test example results





Sampling distribution of the mean estimated using 100,000 Monte Carlo trials.





Scatter diagram for estimated μ and σ. Line is the acceptance limit

$$A=\mu\text{-}k\sigma$$

Empirical test summary results

α and β estimated using 100K Monte Carlo trials for plans with $(p_0,\alpha)=(0.005,0.2)$, $(p_1,\beta)=(0.001,0.1)$

Results for $x_{min}=1000$

Variable	n	k	α	β	n_v/n_a
Exponential(μ)	2	2.43×10^{-3}	0.200	0.082	0.003
Normal($\mu,\sigma=100$)	18	2.886	0.191	0.097	0.023
Normal(μ, σ)	88	2.886	0.191	0.097	0.099
Gamma(10,338 ,δ)	206	2.131	0.193	0.096	0.224
Weibull(10,1995 ,δ)	91	3.623	0.189	0.079	0.117
IG(1502, 100000,δ)	18	2.886	0.173	0.382	unusable

Empirical test results

Results for x_{max}=10,000

Variable	n	k	α	β	n_v/n_a
Exponential(μ)	66	6.26922	0.200	0.082	0.085
Gamma(10,441 ,δ)	77	3.667	0.189	0.104	0.099
Weibull(10,3800 ,δ)	156	3.623	0.188	0.081	0.201

Results for discrete

Variable	n	c	α	β	n_v/n_a
Binomial(n,p)	777	1	0.188	0.100	1
Poisson(n,p)	21	88	0.191	0.097	0.035

ASV fundamental assumption

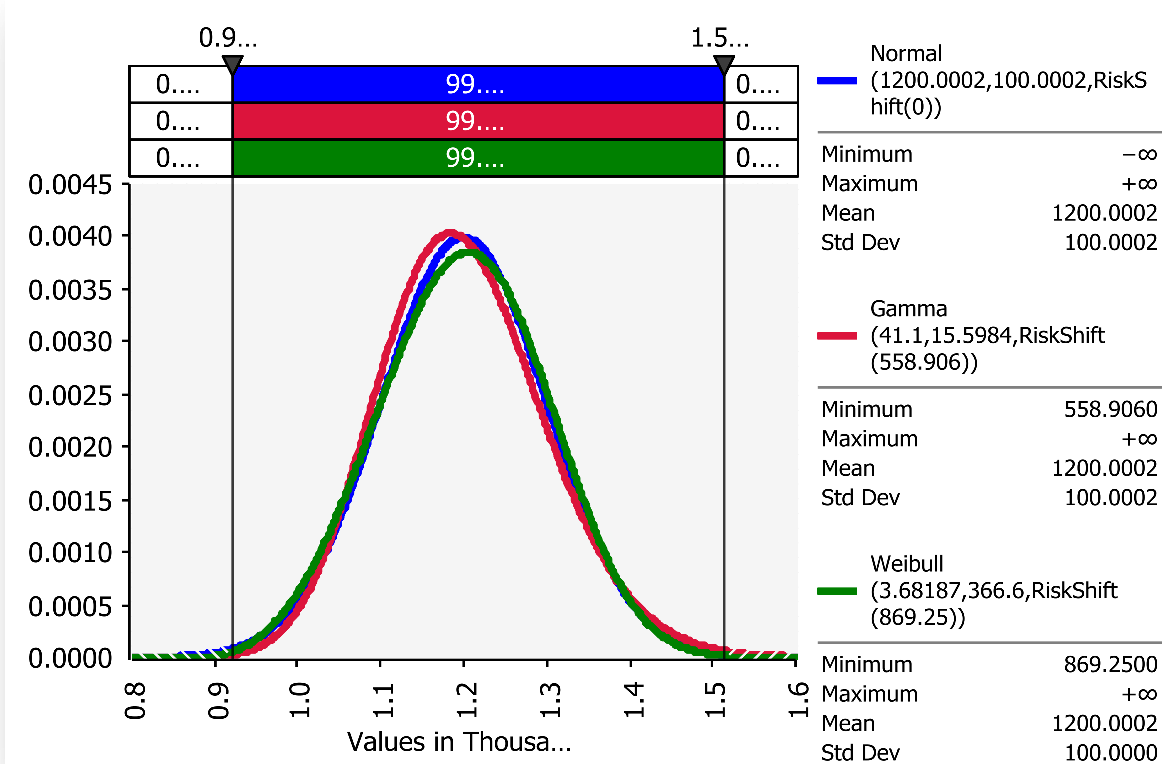
The fundamental assumption of ASV is that the **form of the output distribution is known** a priori. (Moreover, validity testing showed that ASV procedures are robust to error in the shift and scale parameters, but not to shape parameters.)

- In general, the assumption is **unsubstantiated** and the form of the distribution must be determined by **fitting** sample data.
- The question naturally arises**,** "How many trials are required in order to fit the correct form of the parent distribution?"
- Specifically, **"Can we obtain a unique fit to the correct parent distribution based on a sample which is approximately the same size as that specified in the corresponding ASV plan?"**
- The literature appears to be essentially silent on this issue.

  For an exception, see C. Liu (1997) *A Comparison Between the Weibull and Lognormal Models Used to Analyse Reliability Data*, Ph.D. Dissertation, University of Nottingham, UK

Test case 1—Near Normal

Three parent output distributions where chosen with **similar shapes** and **identical means and standard** deviations.



Fitting tests were performed for a requirement with

condition: limit (upper or lower)

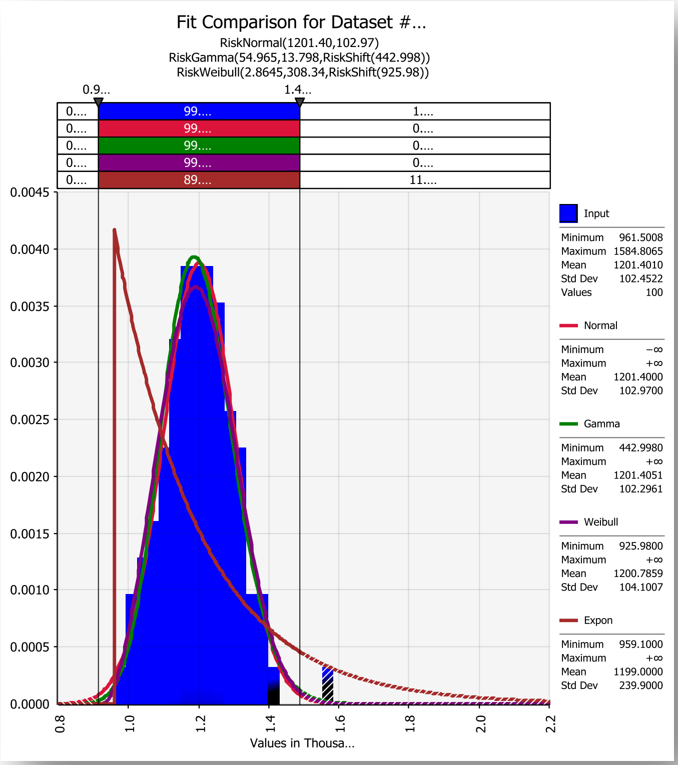
reliability: ρ=0.9973

consumer's risk: β=0.1

(This seems typical of what we had been seeing as Cx level 2 requirements.)

Test 1 procedure and results

- 30 samples of 100 trials each were drawn from the parent Normal distribution.
- Normal, Gamma, Weibull, and Exponential distributions were fit to sample using commercial software (@RISK).
- In general, good fits to the Normal data were achieved with a Normal distribution.
- But good fits also were achieved with Weibull and often Gamma (but not Exponential).
- Fits were compared using three alternative goodness-of-fit (GOF) tests--the “best fit” was sensitive to the GOF test used. (Note: Anderson-Darling is the preferred test here because it gives more weight to the tails.)



- The test procedure was repeated for samples of 300 trials each, with no appreciable change in the nature of the results.
- The test procedure was repeated for samples of 100 and 300 drawn from the Gamma and Weibull parent distributions, with no appreciable change in the nature of the results.

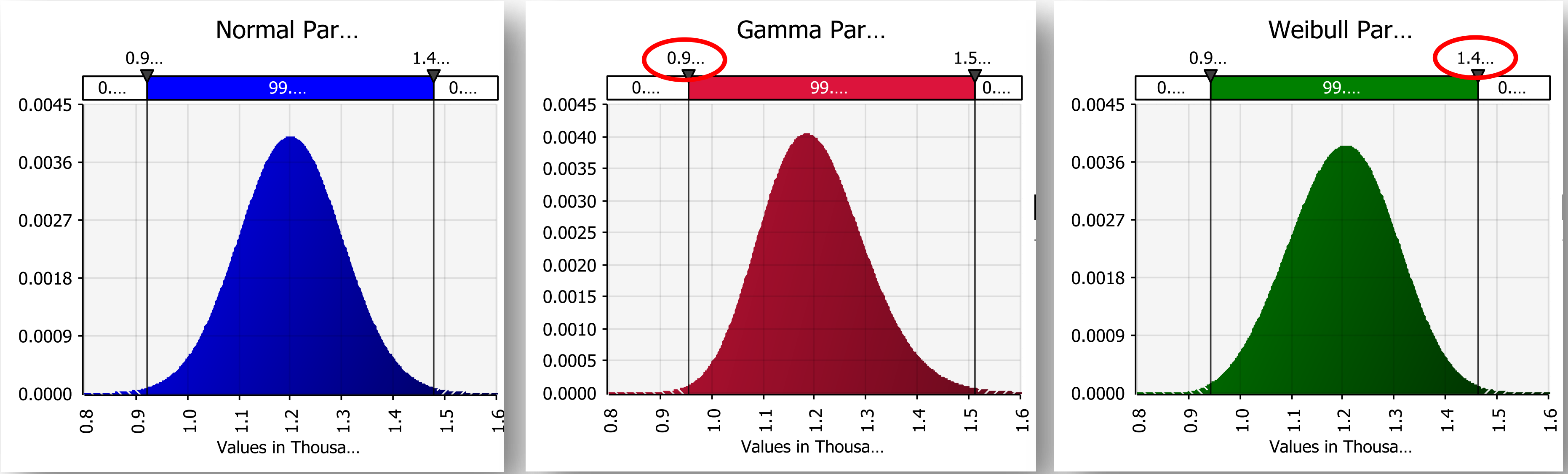
<u>Conclusion:</u> For our example, the size of the sampling plan is inadequate to distinguish the parent distribution for data sets with near-normal shape.

Useful result

Acceptance limits for ρ=0.9973 for the three parent distributions.

Parent	Lower	Upper
Normal	922	1478
Gamma	957	1513
Weibull	943	1463

- The test results were not unexpected—these distributions are very similar in shape *overall*.
- Our interest is in the (small) differences in the tails of these of these distributions.
- Note that the distribution with the **smallest *k* factor provides the greatest protection against accepting a nonconforming design** (i.e., the largest lower limit and the smallest upper limit).



Application

Upper limit sampling plans (α=0.2)

Variable	*n*	*c*	*k*	*A*
Binomial (ASA plan)	2959	4		
Normal(1200, 0)	257		2.968	1497
Gamma(41.1, 15.5984, 558.906)	224		3.378	1538
Weibull(3.68187, 366.6, 869.25)	296		2.787	1479

Lower limit sampling plans (α=0.2)

Variable	*n*	*c*	*k*	*A*
Binomial (ASA plan)	2959	4		
Normal(1200, 0)	257		2.968	903
Gamma(41.1, 15.5984, 558.906)	353		2.566	943
Weibull(3.68187, 366.6, 869.25)	615		2.677	932

The **plan with the tightest bound will yield the most conservative decision**—one that guarantees the consumer's risk is no greater than specified.

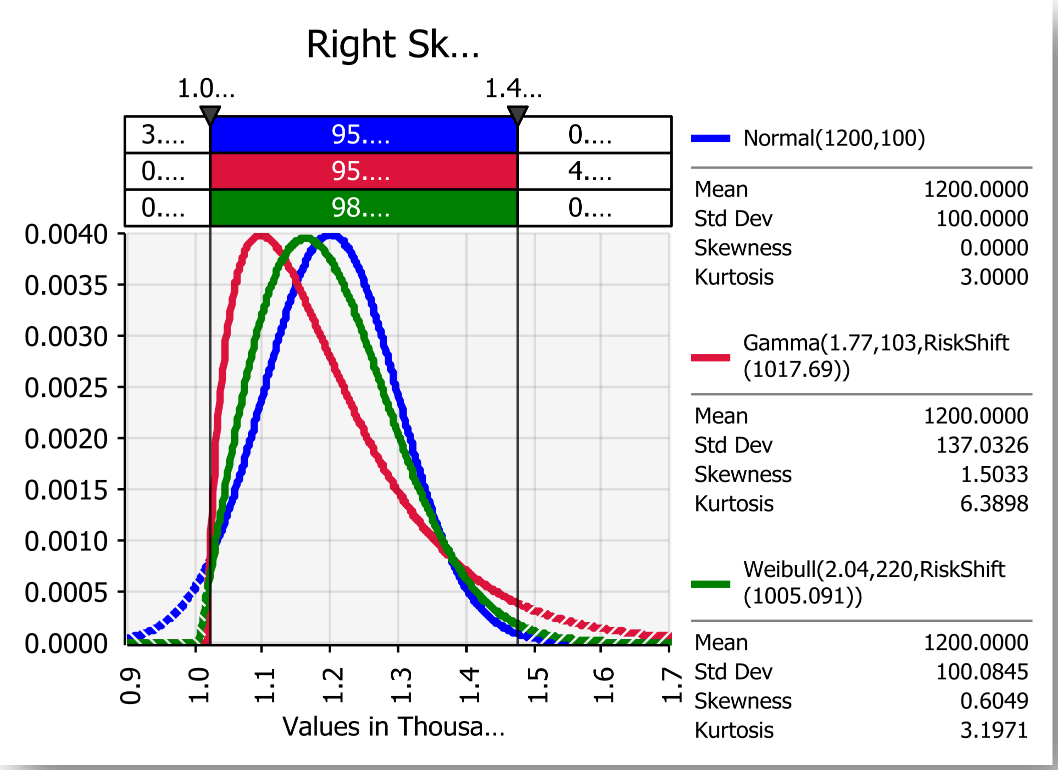
For the example requirement, the Weibull plan provides the tightest (least) upper bound.

The Gamma plan provides the tightest (greatest) lower bound.

These plans provide an order-of-magnitude reduction in computing effort.

Test case 2—Right skew

The second case considers Weibull and Gamma distributions with identical means and moderate skew.



Fitting tests were performed for the same requirement

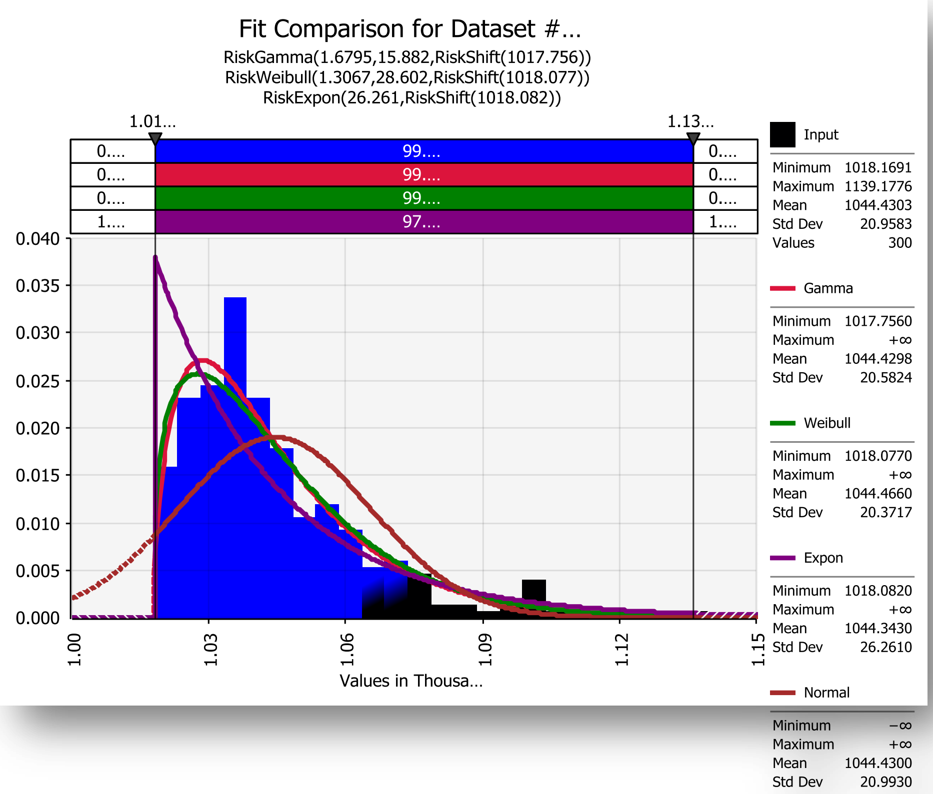
condition: limit (upper or lower)

reliability: ρ=0.9973

consumer's risk: β=0.1

Test 2 procedure and results

- 30 samples of 300 trials each were drawn from the parent Weibull distribution.
- Normal, Gamma, Weibull, and Exponential distributions were fit to sample using commercial software.
- In general, good fits to the Weibull data were achieved with the Weibull distribution and sometimes the Gamma distribution.
- The skew is sufficiently large that the data are not mistaken as Normal.
- The skew is sufficiently small that the data are not mistaken as Exponential.
- The test procedure was repeated for a the parent Gamma distribution, with no appreciable change in the nature of the results.



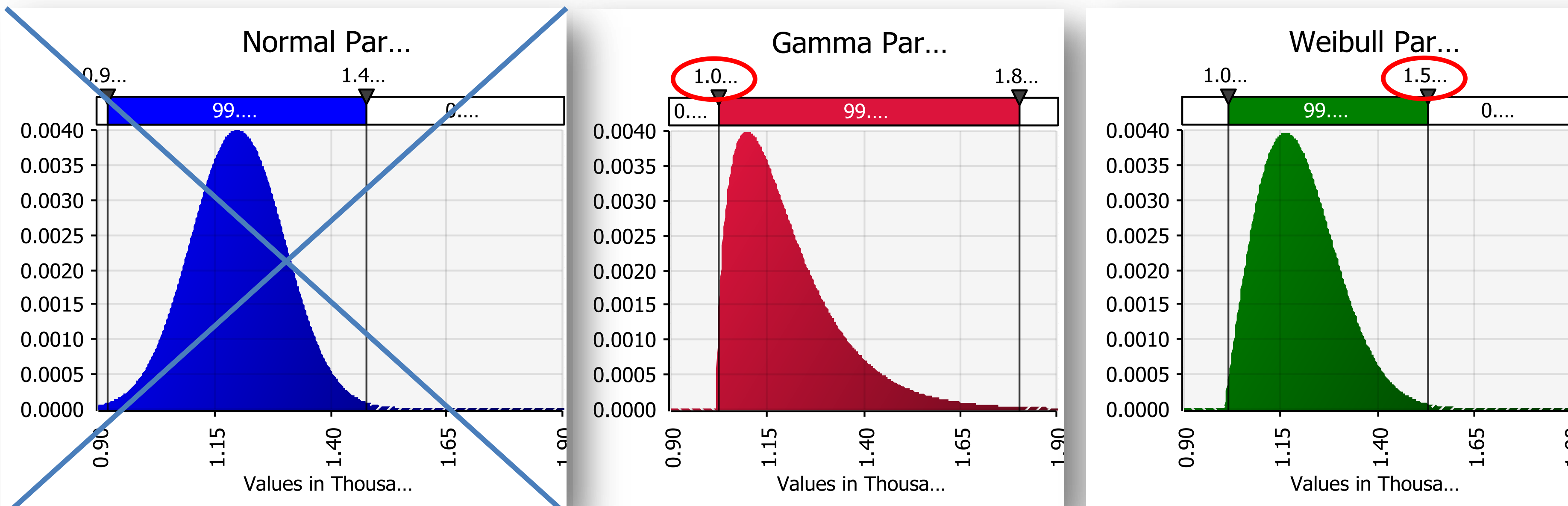
Conclusion: For our example, the size of the sampling plan is inadequate to distinguish between Weibull and Gamma, but adequate to rule out Normal and Exponential parents.

Useful result 2

Acceptance limits for ρ=0.9973 for the three parent distributions.

Parent	Lower	Upper
Normal	922	1478
Gamma	1023	1806
Weibull	1017	1531

- As before, the test results were not unexpected and our interest is in the (small) differences in the tails of these of these distributions.
- As before, the distribution with the **fattest tail in the direction of the limit provides the greatest protection** against accepting a nonconforming design.
- However, with sufficient skew as in this example, we can rule out Normal as the parent (poor fit).



Application

Upper limit sampling plans (α=0.2)

Variable	*n*	*c*	*k*	*A*
Binomial (ASA plan)	2959	4		
~~Normal(1200, 0)~~	~~257~~		~~2.968~~	~~1497~~
Gamma(1.77,15.5984,1017.7)	278		4.915	1873
Weibull(2.04,220,1005.9)	264		3.557	1557

Lower limit sampling plans (α=0.2)

Variable	*n*	*c*	*k*	*A*
Binomial (ASA plan)	**2959**	4		
~~Normal(1200, 0)~~	~~257~~		~~2.968~~	~~903~~
Gamma(1.77.15.5984,1017.7)	**25009**		1.304	1018
Weibull(2.04,220,1005.6)	3651		1.855	1015

Once again, the plan with the tightest bound will yield a conservative decision.

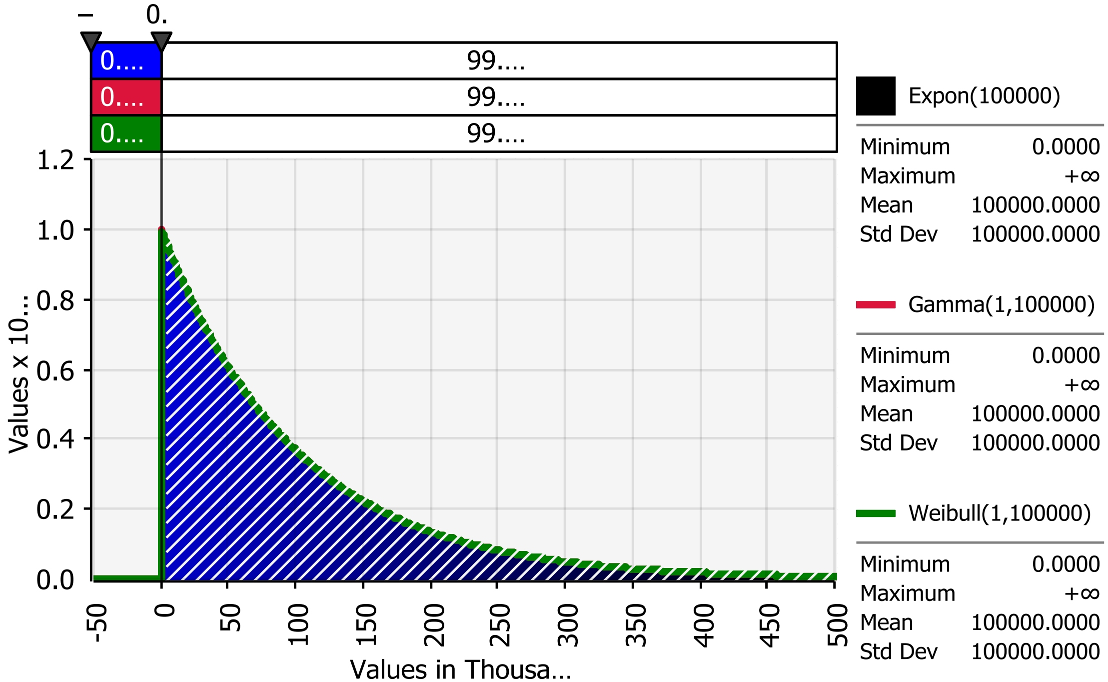
The Weibull plan provides the tightest (least) upper bound and can be used in this case.

The Gamma plan provides the tightest (greatest) upper bound, but is very large. Obviously, the ASA plan is preferable in this case.

The UL plan provides an order-of-magnitude reduction in computing effort.

Test case 3—Near Exponential LL

- A typical application is lifetime data, where the random variable X represents the time at which a component fails and the condition is the lower limit $L = X_{\min}$.
- Note that Expo(θ), Gamma(1, θ), and Weibull(1, θ) are the *same* distribution.



Plan comparisons

- The Exponential plan is *very* small—more than three orders of magnitude smaller than the attributes plan.
- The Gamma and Weibull plans are identical and *very* large—almost three orders of magnitude larger than the attributes plan.

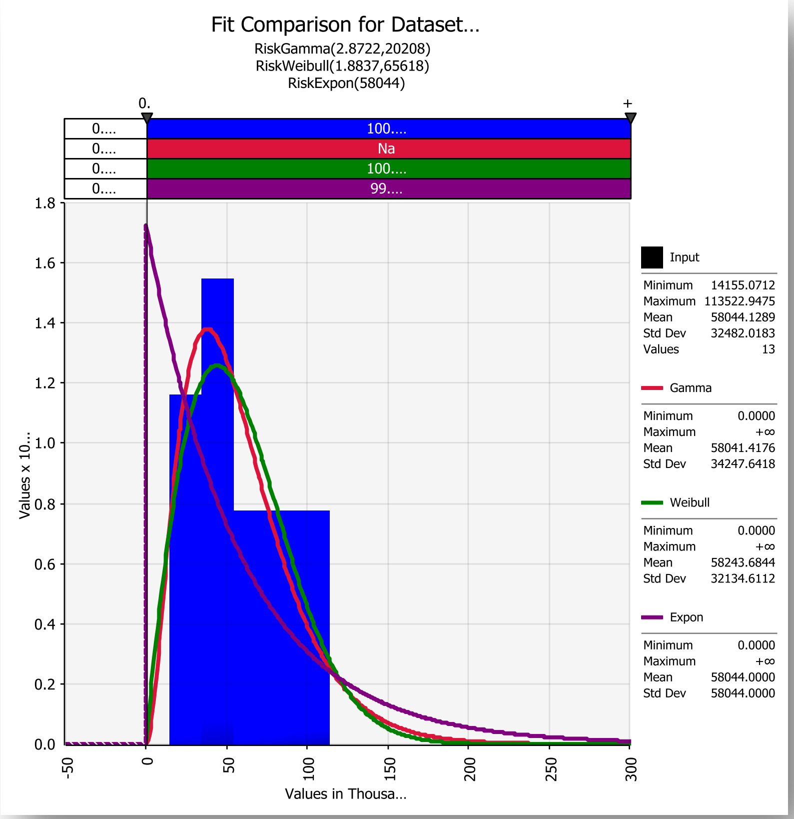
  This remarkable difference in *n* flows from the fact that Exponential has a single parameter and fixed shape–if we *know* a prior the parent is exponential, then we need only estimate the mean.
- The Gamma and Weibull plans are the same, both derived from the same approximation. These are more conservative than the exponential (the true value of A=270.37), but unusable in this application because of their size.

Lower limit sampling plans (α=0.02)

Variable	*n*	*c*	*k*	*A*
Binomial (ASA plan)	6580	12		
Exponential(100000)	13		0.9806	1940.00
Gamma(1, 100000)	3819299		0.9979	2060.45
Weibull(1,100000)	3819299		0.9979	2060.45

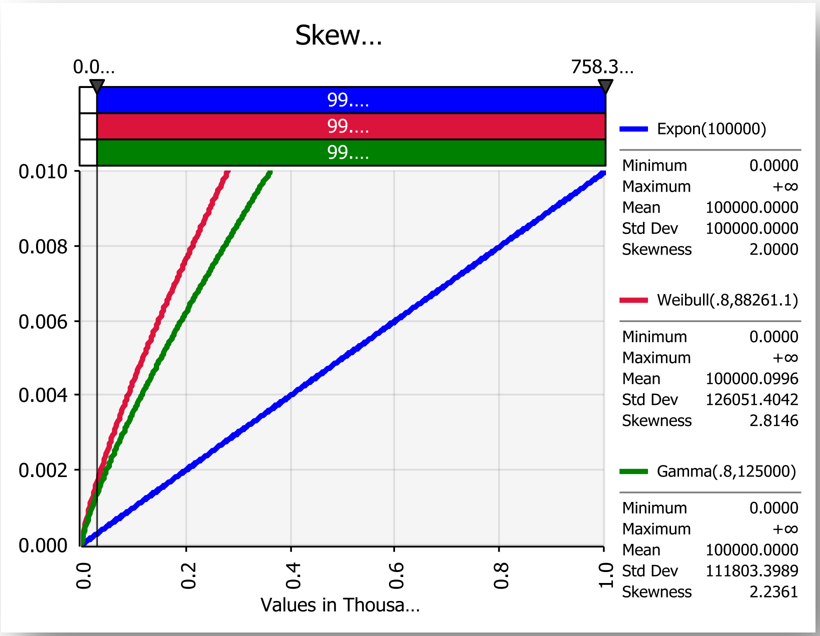
Test case 3—modest skew

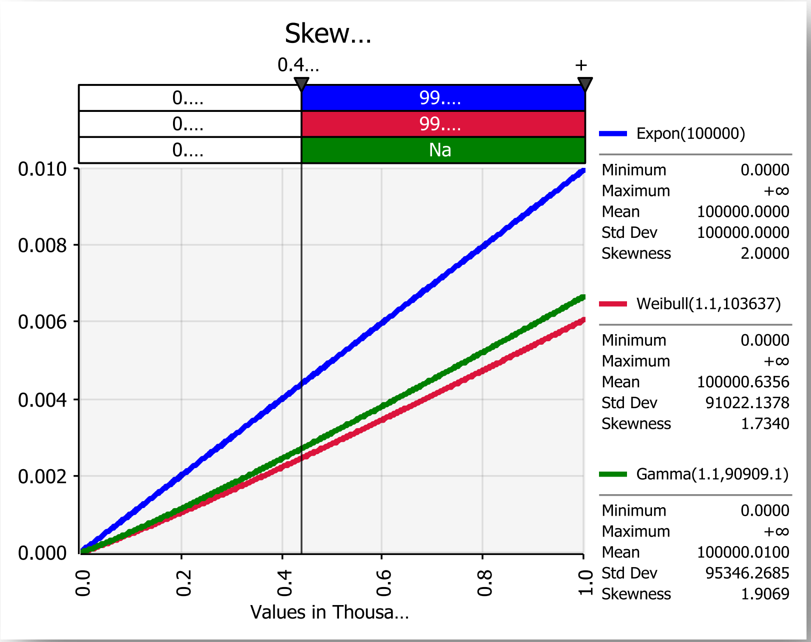
- 30 samples of 13 trials each were drawn from the parent Exponential distribution
- Gamma, Weibull, and Exponential distributions were fit to these datasets (shifts set to zero for lifetime data).
- Fits were compared using three alternative goodness-of-fit tests.
- Best fits were dependent GOF test and there were many ties for best fit.
- Acceptable fits to exponential (*p*-value$\geq$0.15) where obtained in most cases (30 Chi-squared, 25 K-S, 23 A-D).
- Exponential or binomial are the only practical plans in this case and acceptable Exponential fits most often can be achieved for Exponential parent distributions.



Effect of skew

But what about datasets from Weibull and Gamma parents masquerading as Exponential?





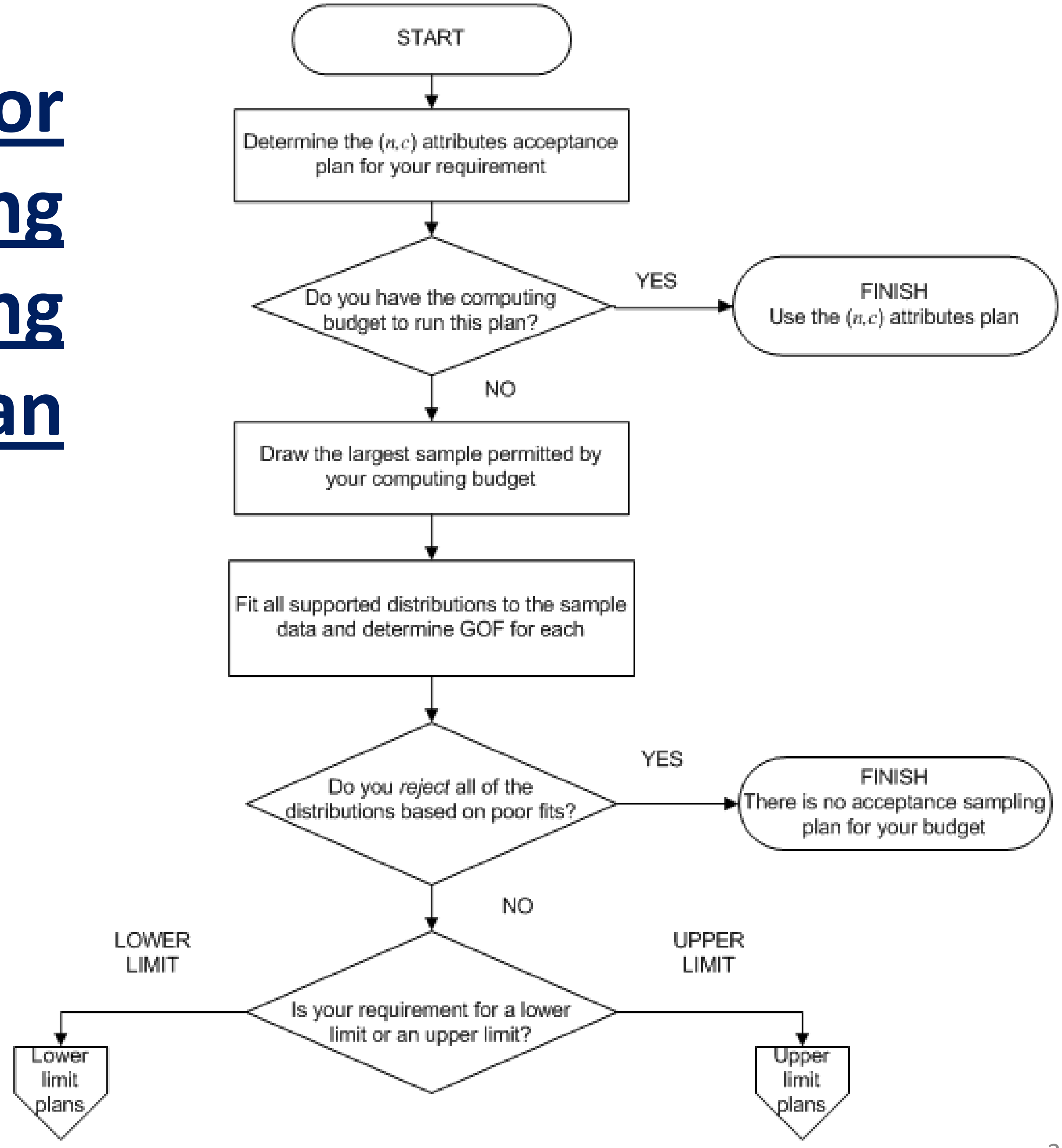
The Exponential plans are conservative for Wiebull parents with skew >2 and non-conservative for skew < 2. Accepting the Exponential fit in the second case will result in a modesty lower reliability than specified (0.995 rather than 0.997 in this example).

Can modest skew be detected?

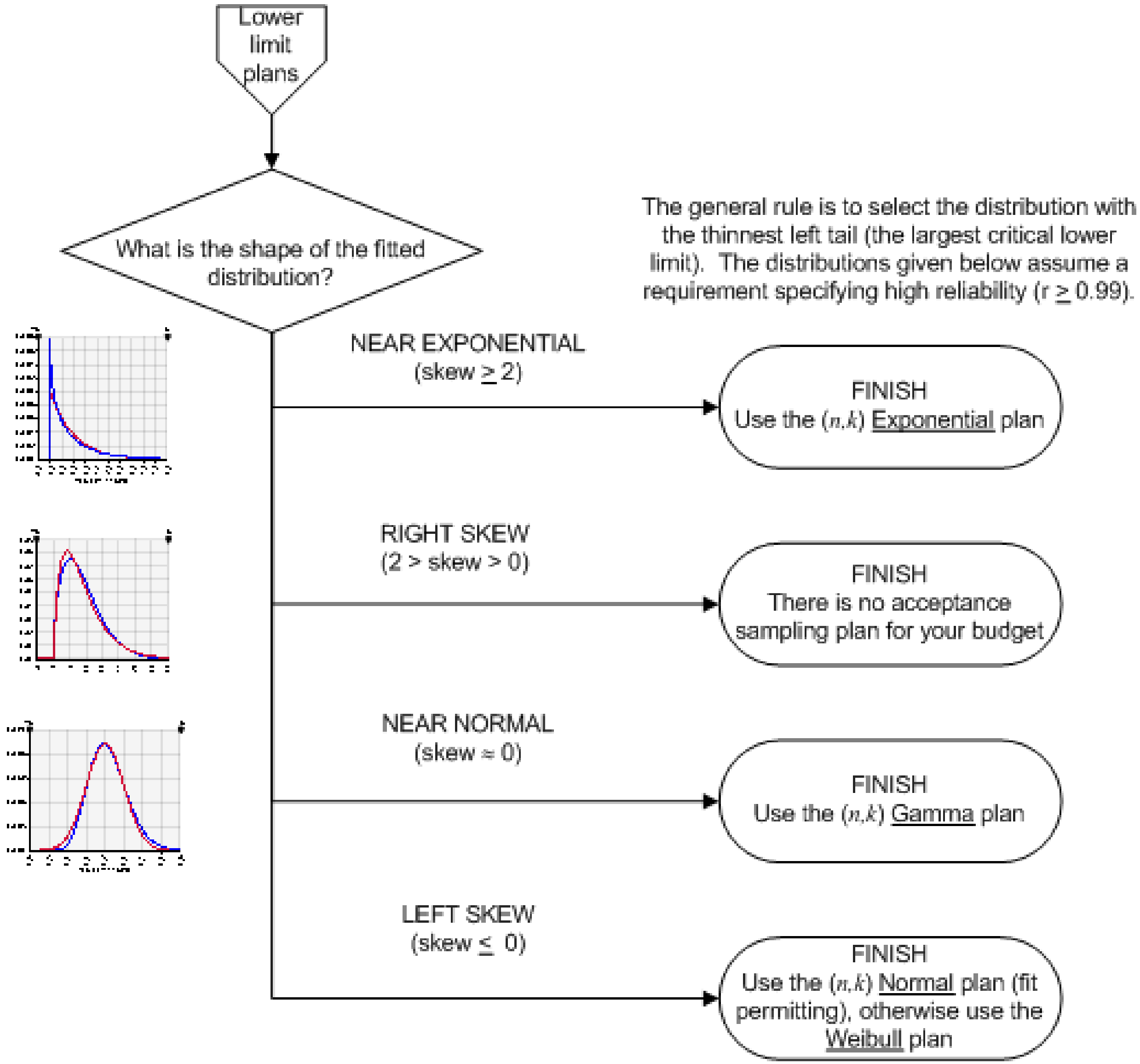
- 30 samples of 13 trials each were drawn from the parent Weibull distribution with skew=0.631
- Gamma, Weibull, and Exponential distributions were fit to these datasets (shifts set to zero for lifetime data).
- A-D and K-S tests typically showed very poor fits to Exponential even with these small samples (Chi-squared test appears to lack power).
- P-P and Q-Q plots illustrated that the Exponential fits were poorest in the region of interest .

Conclusion: With a small sample and skew $\approx$ 2 it is not possible to discern the parent distribution; with lesser skew Weibull and Gamma will not be confused with Exponential.

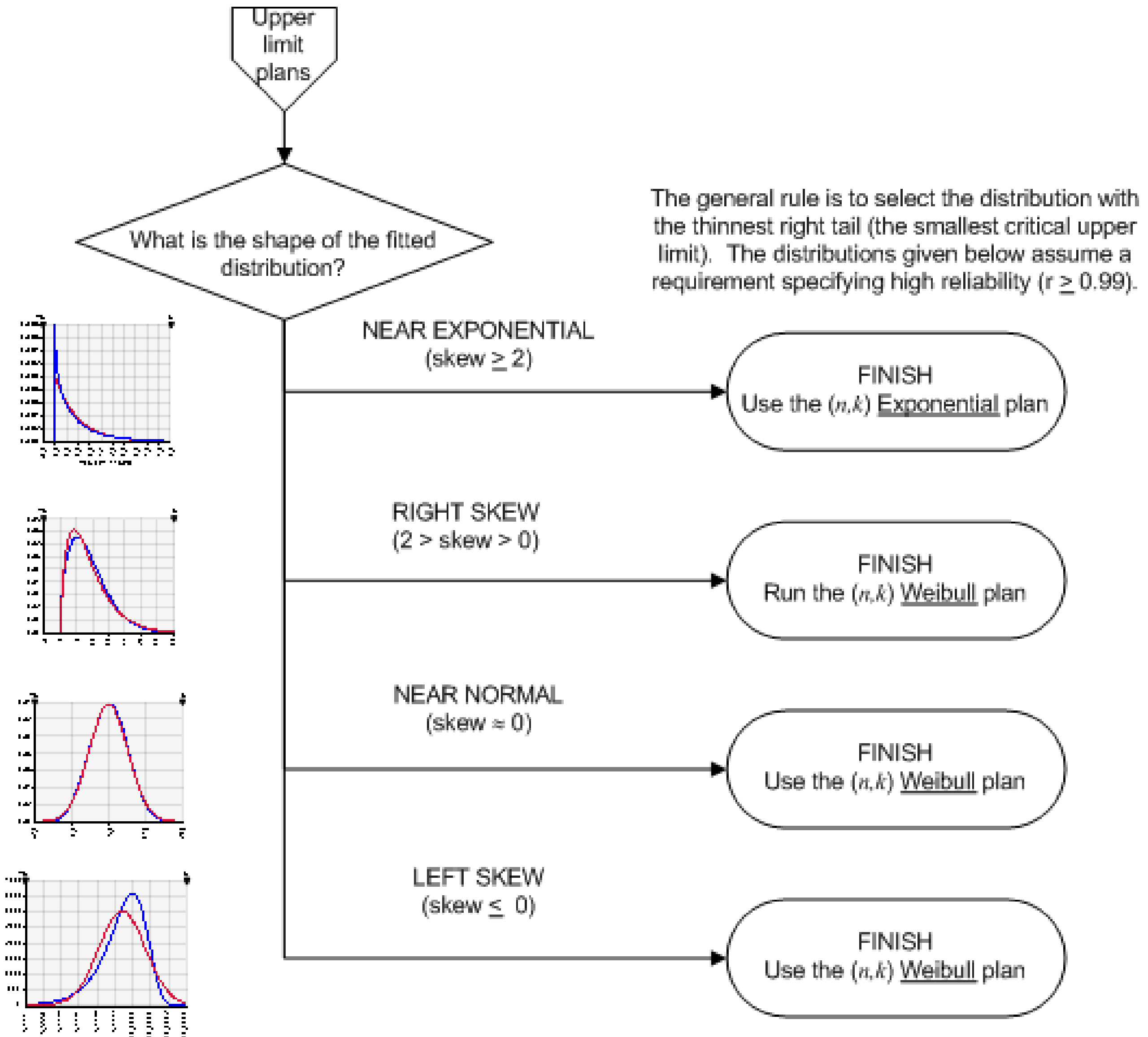
Procedure for determining a sampling plan



Procedure (lower limit plans)



Procedure (upper limit plans)



Summary

- ASA plans are preferred when computational demands can be met. These plans are exact and transparent.
- ASV plans are a viable alternative, when ASA plans are too large. These plans are inherently approximate; the data (perhaps transformed) must fit the a distribution for which ASA calculator is available. The assistance of a statistician would be beneficial.
- For data with skew less than Exponential, the Normal, Gamma, or Weibull plan with the tightest bound is a good choice—it is conservative and can provide an order of magnitude reduction in computational effort.
- For near-Exponential data with a lower limit, the Exponential plan is not necessarily conservative (and should be applied intelligently and with caution)—but can provide several orders of magnitude reduction in computational effort.
- In some applications, uncertainties in the protection afforded by Exponential may be inconsequential when model error is considered.
- When applying the Exponential, a good practice is to make as many trials as feasible and then attempt a fit to the distributions currently supported.

Contributions

Variables acceptance can reduce sample sizes and the reduction can be as much as one, two, and even three orders of magnitude depending on the distribution and OC. But this isn't always the case. Gamma and Weibull plans become larger than attributes plans as the shape parameter decreases.

Normal plans don't work well for inverse Gaussian. But the published inverse Gaussian plans don't work either. We've found the error in the derivation and I think I may have a fix given the time to mess with it. That's news.

The purpose of variables acceptance is to reduce sample size. But in accomplishing this we can't be sure that we satisfy the fundamental assumption that the distribution is known, at least for the OC we are interested in. Assuming we want to be conservative with respect to consumer's risk, we've developed a procedure to overcome this issue.

Statistical Tolerance Bounds: Overview and Applications to Space Systems

James Womack
The Aerospace Corporation

3 May 2011

Statistical Intervals

- Decisions are frequently made based on limited sample data
  - *Examples: Strength of a material must be at least X*

    Minimum strength of a material must exceed max load applied

    Determine minimum strength of a material over several environments
- Sample data can be used to estimate mean strength or probability of exceeding limits, but provides no information about the precision of the estimates
  - *There may be big differences between the estimated values and what the true values are, if unlimited data were available*
- Statistical intervals quantify the uncertainty associated with an estimate

James.M.Womack@aero.org

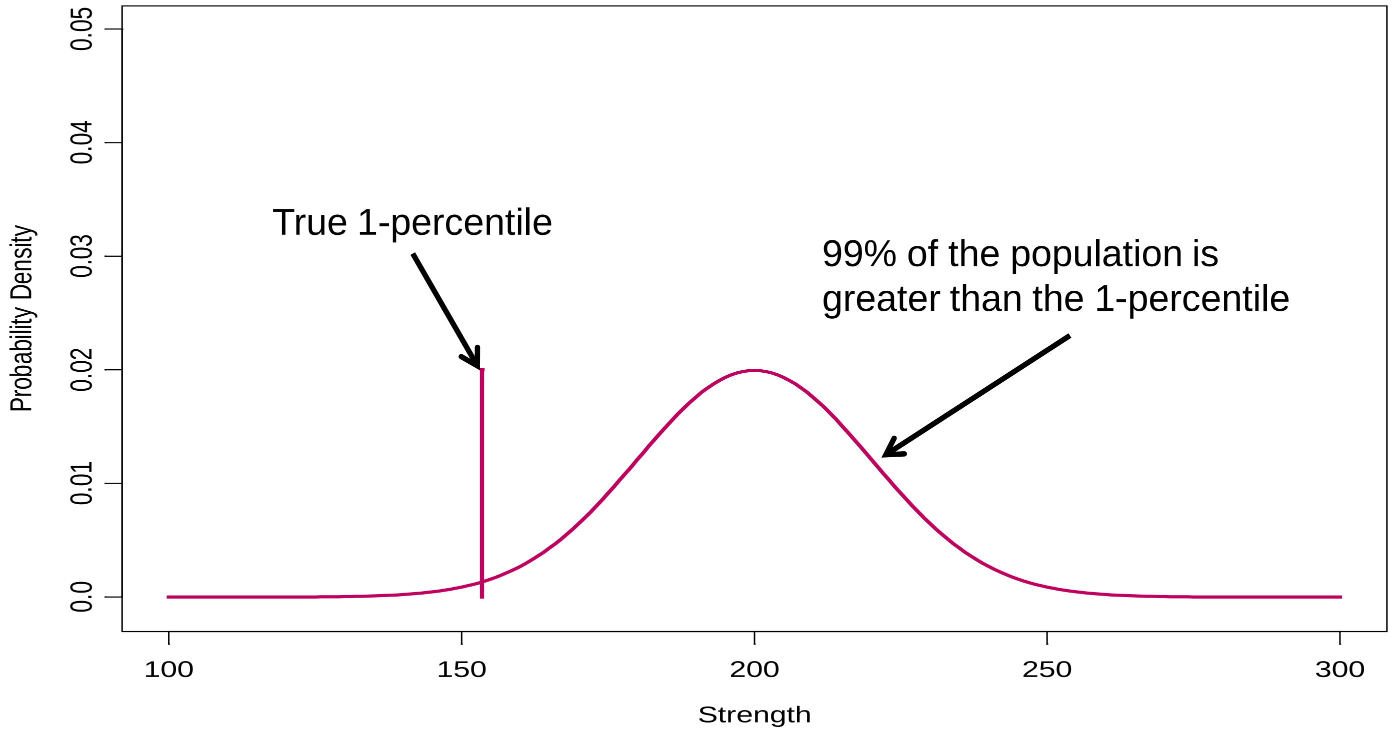
Types of Statistical Intervals

- Appropriate type statistical interval depends on the application
  - *Confidence and Tolerance Intervals are for describing the population or process from which the sample data has been selected*
  - *Prediction Intervals are for predicting the results of a future sample*
- Confidence intervals: enclose means, variances and other population parameters
- Tolerance intervals are used to contain a specified proportion of a population
  - *Lower tolerance bound for bounding the population from below (minimum strength)*
  - *Upper tolerance bound for bounding the population from above (maximum load)*
  - *Tolerance interval for enclosing the population both above and below*
- Example: If T is an upper tolerance bound with 90% confidence for .99 of the population, then we are 90% confident that 99% of the population is less than T
  - *T is denoted as a (.99, .90), where 99 represents the proportion of the population bounded and 90 represents the confidence level*

James.M.Womack@aero.org

1-Percentile of Normal Distribution

- Assume the population is normal
  - *Mean (μ = 200) and variance (σ^2 = 400)*
- Then the 1-percentile = $\mu - 2.326\sigma = 153.5$



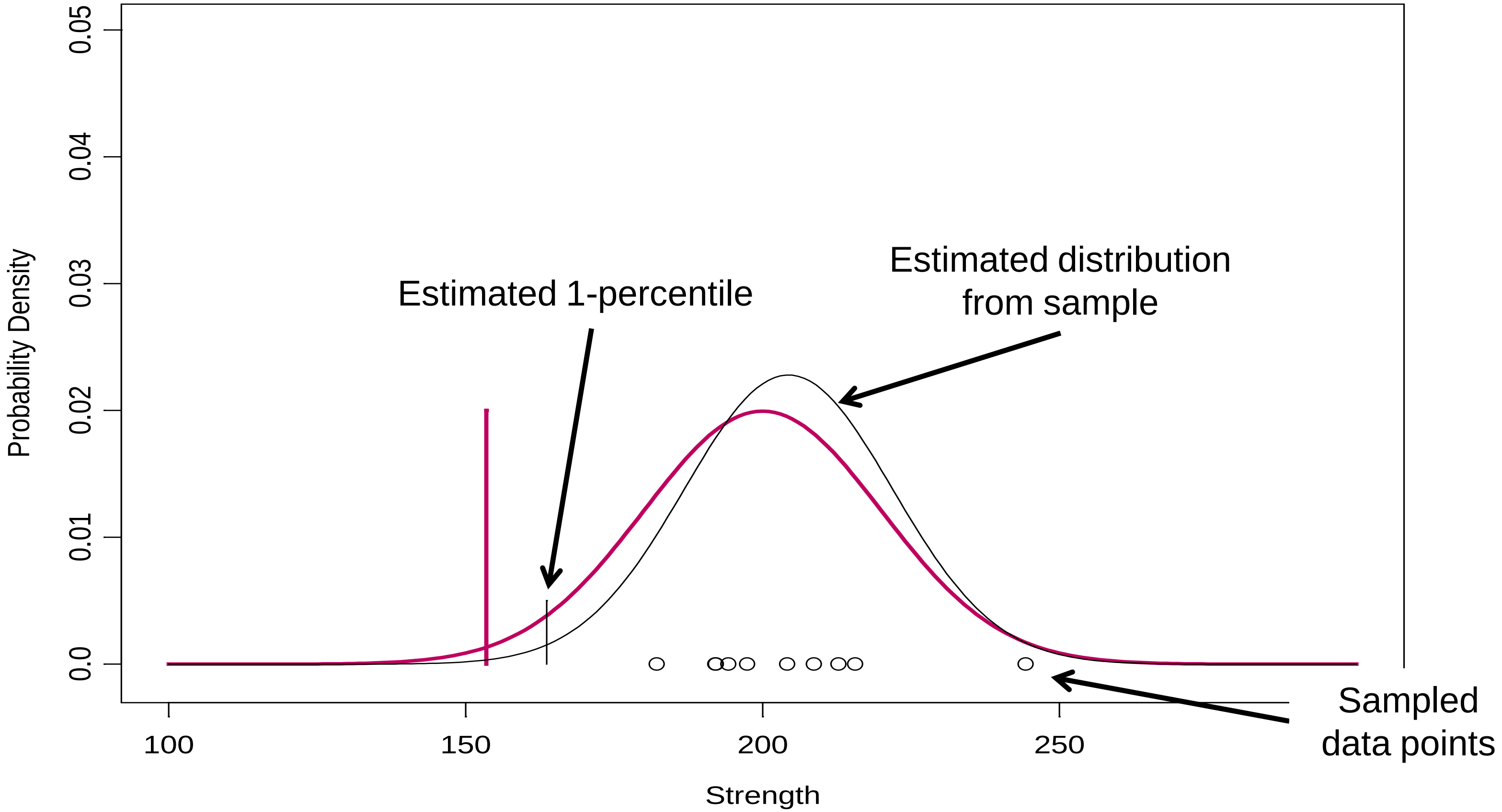
- When the mean and variance are both unknown, we must sample from this distribution to estimate the 1-percentile

James.M.Womack@aero.org

AEROSPACE

Sample of 10 Measurements Used to Estimate 1-Percentile

- Use sample mean, $\bar{x}$, in place of μ and sample variance, s^2, in place of σ^2
- Estimated 1-percentile = $\bar{x} - 2.326s$

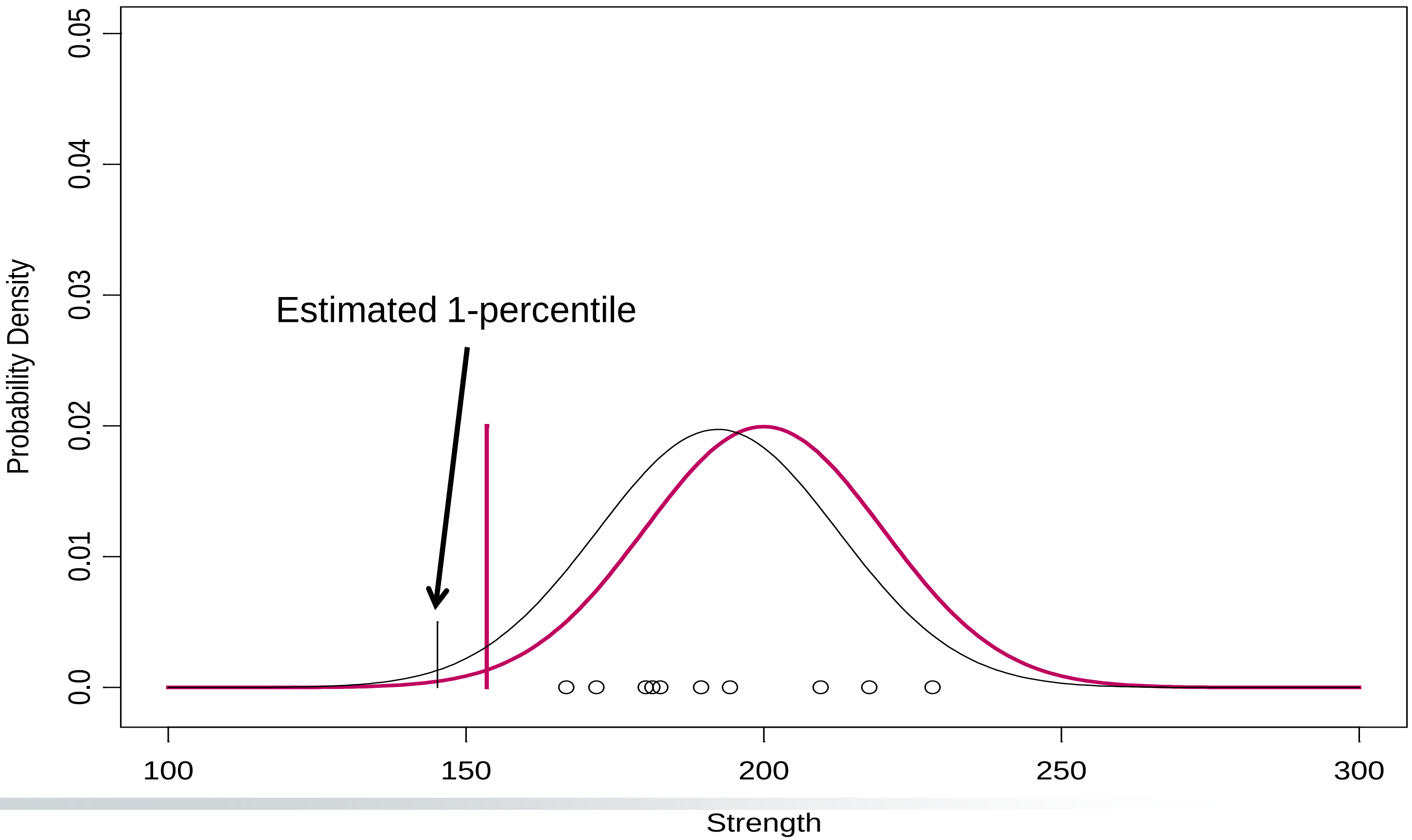


James.M.Womack@aero.org

AEROSPACE

Another Sample of 10 Measurements Used to Estimate 1-Percentile

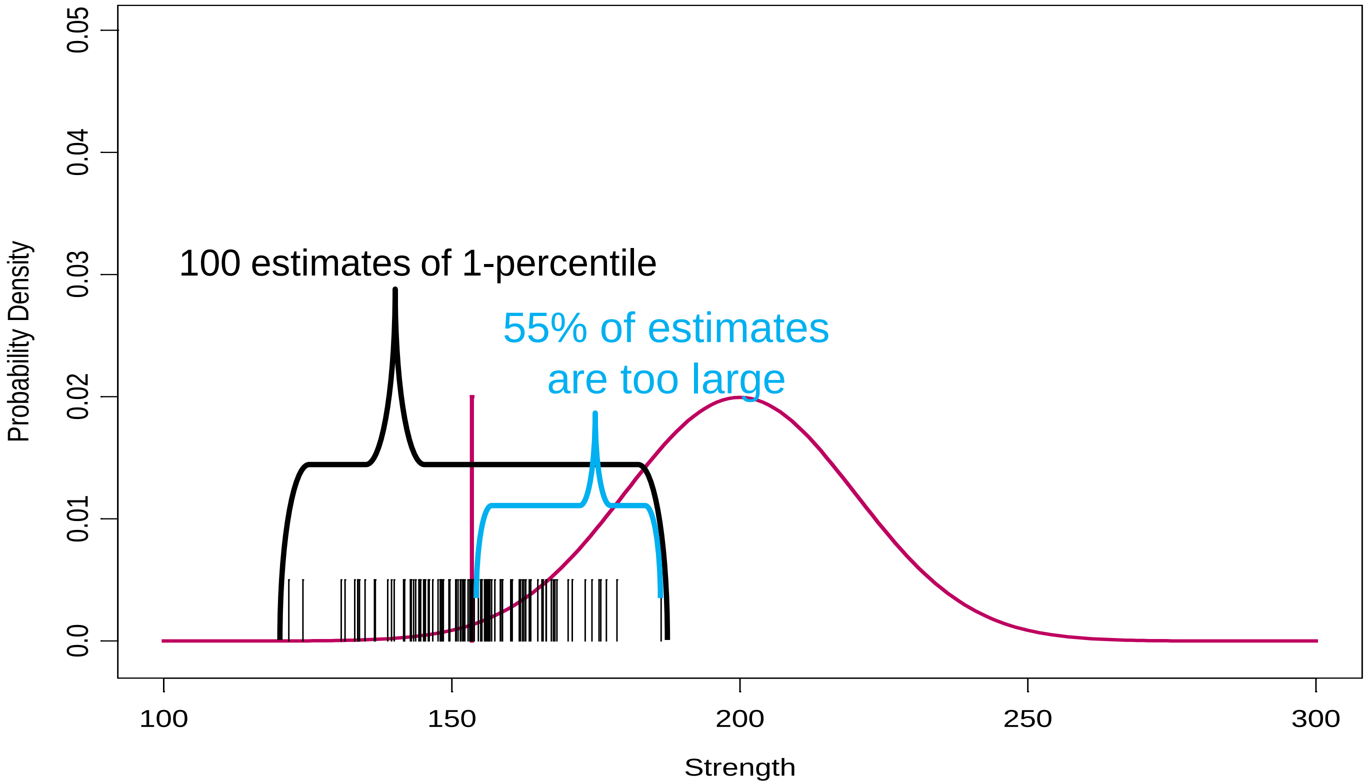
- Use sample mean, $\bar{x}$, in place of μ and sample variance, s^2, in place of σ^2
- Estimated 1-percentile = $\bar{x} - 2.326s$



James.M.Womack@aero.org

100 Samples of 10 Measurements Used to Estimate 99-percentile

- Use sample mean, $\bar{x}$, in place of μ and sample variance, s^2, in place of σ^2
- Estimated 1-percentile = $\bar{x} - 2.326s$

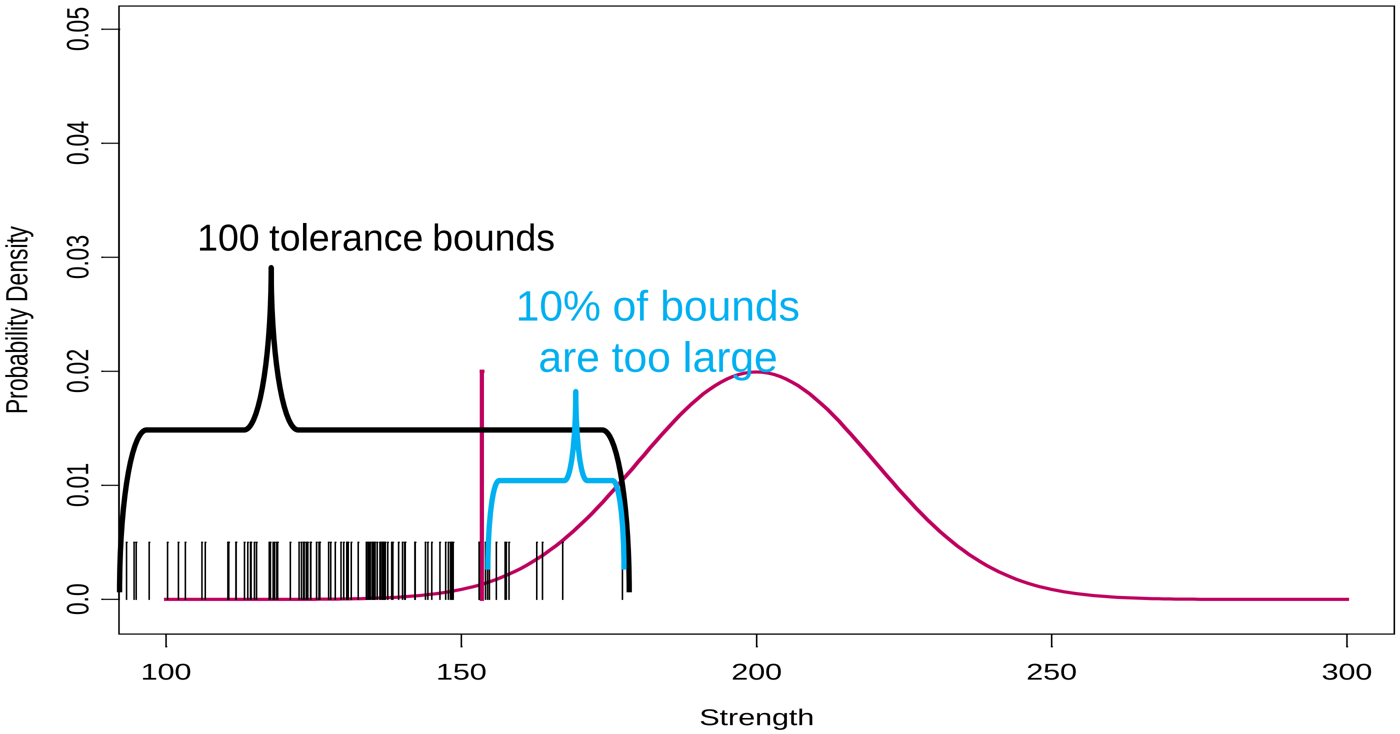


- A (.99,.90) lower tolerance bound is of the form: $\bar{x} - ks$
  - *k is selected so that 90% of the samples will produce a bound that encloses 99 percent of the population (k = 3.532 for sample size of 10)*

James.M.Womack@aero.org

(.01,.90) Lower Tolerance Bounds: 100 Samples of 10 Measurements

- Use sample mean, $\bar{x}$, in place of μ and sample variance, s^2, in place of σ^2
- (.99,.90) Lower tolerance bound = $\bar{x} - 3.532s$



- Increasing sample size will decrease spread of tolerance bounds
- Increasing confidence level will decrease percent of time tolerance bound is too large

James.M.Womack@aero.org

Tolerance Bounds for Normal Distribution

- $X_1, X_2, \ldots, X_n$ sample from normal population, $N(\mu,\sigma^2)$, with unknown μ and σ^2
  - *Let z_p be the p quantile of standard normal distribution*
  - *p quantile of $N(\mu,\sigma^2)$ is $q_p = \mu + z_p \sigma$*
- The 1-α upper confidence bound for q_p is a (p,1- α) one-sided upper tolerance bound for $N(\mu,\sigma^2)$
  - *Find k such that* $P_{\overline{X},S}\left\{P\left(X < \overline{X} + kS \mid \overline{X}, S\right) > p\right\} = 1-\alpha$
  - *Or* $P\left\{\dfrac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} + z_p\sqrt{n}}{\dfrac{S}{\sigma}} > -k\sqrt{n}\right\} = 1-\alpha$ *(Non-central t distribution)*
  - *k =* $\dfrac{1}{\sqrt{n}} t_{n-1;1-\alpha}(z_p\sqrt{n})$
- The (p,1- α) upper tolerance bound is $\overline{X} + t_{n-1;1-\alpha}(z_p\sqrt{n})\dfrac{S}{\sqrt{n}}$
- The (p,1- α) lower tolerance bound is $\overline{X} - t_{n-1;1-\alpha}(z_p\sqrt{n})\dfrac{S}{\sqrt{n}}$

James.M.Womack@aero.org

Splus Code for Normal Tolerance Bound

```
# Computes (p,1-alpha) upper tolerance bound on X ~ Normal
#
# Inputs: xdata (vector of sample data), p = percentile, alpha = 1 - confidence level

library(envstats)            # library for non-central t

UpperToleranceBound <- function(p,alpha,xdata){
        n <- length(xdata)
        Ssq <- var(xdata)
        UTB <- mean(xdata) + qt(1-alpha, n-1, ncp = qnorm(p)*sqrt(n))*sqrt(Ssq/n)
        UTB
}
```

```
# Example
> xdata <- c(1.822938, 1.143871, 0.972309, -0.078231, 0.480773, 0.710025, -0.573717,
0.272126, 0.016359, -0.596675)
> UpperToleranceBound(.99,.10,xdata)
[1] 3.1371
```

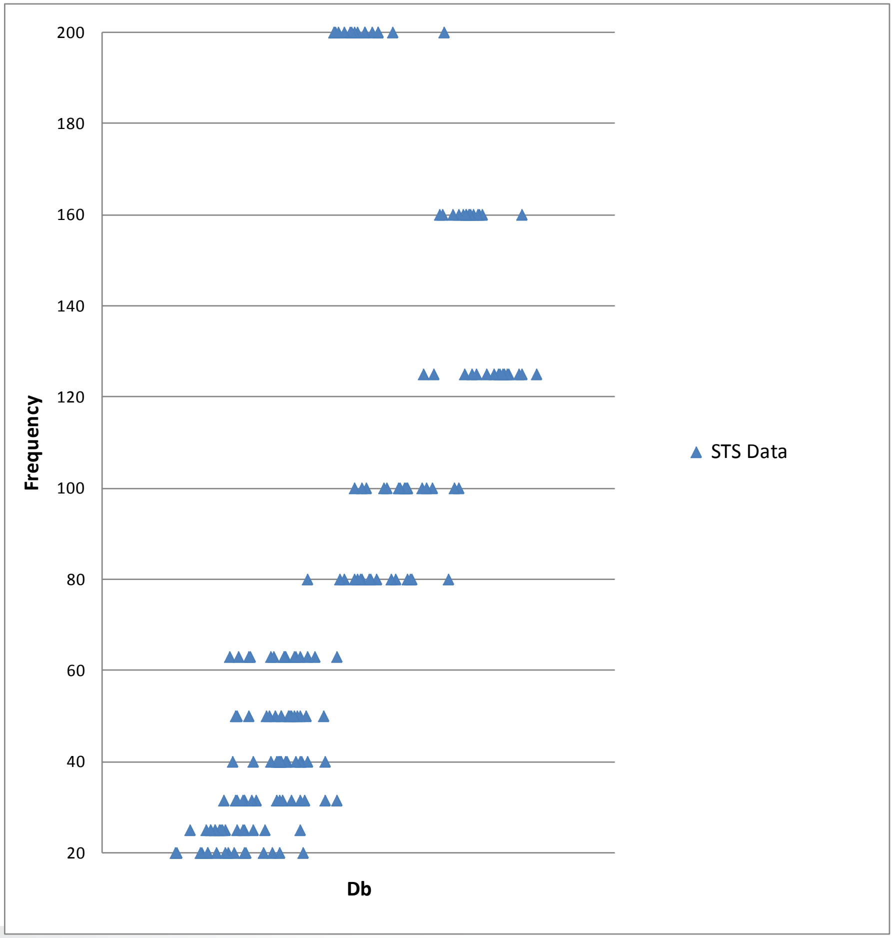

James.M.Womack@aero.org

Orbiter Heat Shield Main Engine Ignition Acoustic Environment

Upper Tolerance Bound Example

- After post flight analysis, 16 flights remaining with data judged suitable for use in updating base heat shield environment limits
- MIL-STD- 1540E and NASA STD –HDBK-7005 approach for this limit:
  - (.99,.90) upper tolerance bound used for qualification testing
  - (.95,.50) upper tolerance bound used for acceptance test

- STS acoustic vibration dataset:



James.M.Womack@aero.org

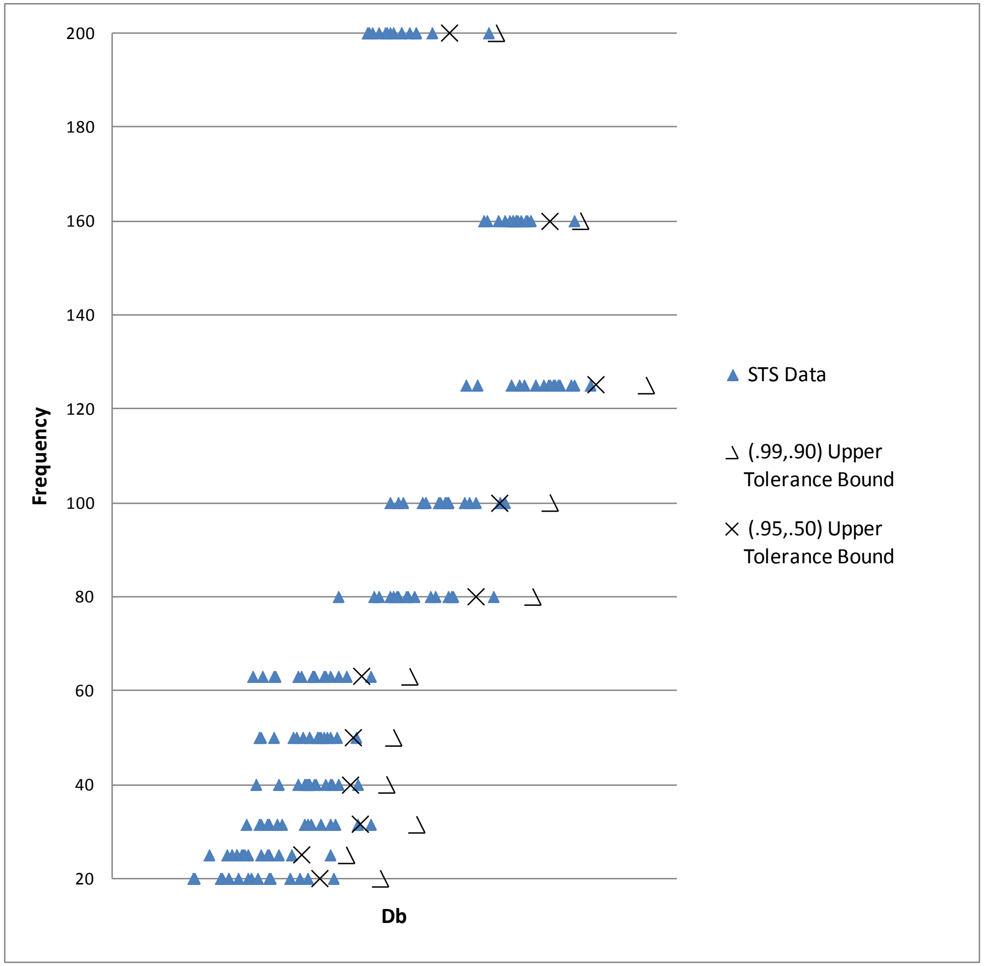
AEROSPACE

Orbiter Heat Shield Main Engine Ignition Acoustic Environment

Upper Tolerance Bound Example (2)

- Assume independence between frequency bands
- Probability plots show data can be reasonably modeled by normal distribution
- Normal upper tolerance bounds:
  - *(.99,.90):* $\bar{x} + 3.172s$
  - *(.95,.50):* $\bar{x} + 1.678s$

- Results



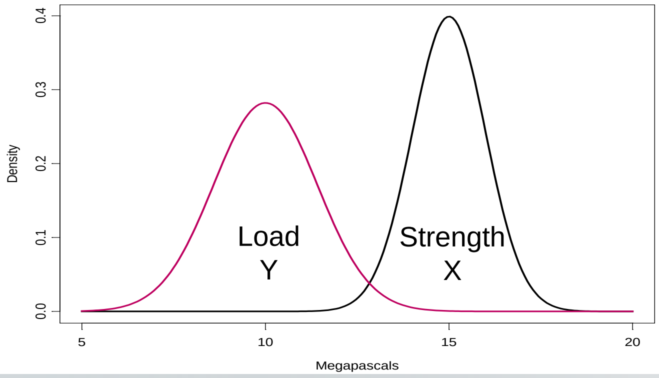
James.M.Womack@aero.org

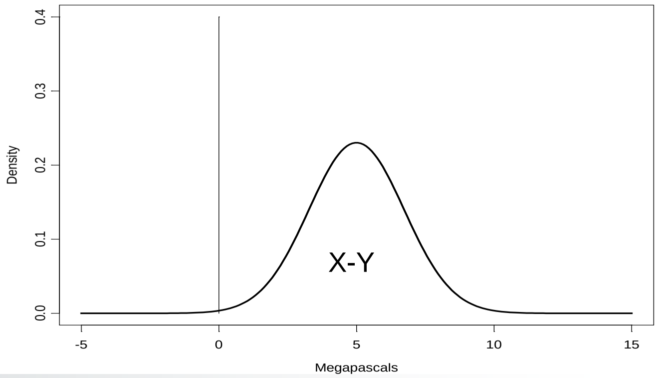
AEROSPACE

CEV Parachute Assembly System (CPAS) Reliability

Load-Strength Reliability Problem

- To calculate the reliability for the main and drogue parachute components, both the strength and load distributions are characterized by independent normal distributions
- Reliability, *R*, is the probability that the strength, *X*, is greater then the load, *Y*, i.e. P(*X*-*Y* > 0)
  - *Distribution of X-Y is normal with mean* μ_X - μ_Y *, and variance* $\sigma_X^2 + \sigma_Y^2$
  - *R* = $1 - \Phi\left(\dfrac{\mu_Y - \mu_X}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right)$*, where* Φ *is the cumulative normal distribution*
  - *Since the means and variances are unknown, we can only estimate the reliability so we also need a lower confidence bound on R*
- If we select a quantile, p, such that the (p,1-α) lower tolerance bound = 0, then p is a 100(1- α)% lower confidence bound on R





James.M.Womack@aero.org

Lower Tolerance Bounds for X-Y

- $X_1, X_2, \ldots X_{n_1}$ sample from normal population, $N(\mu_X, \sigma_X^2)$,
- $Y_1, Y_2, \ldots Y_{n_2}$ sample from normal population, $N(\mu_Y, \sigma_Y^2)$,
- 1-p quantile is: $L_p = \mu_X - \mu_Y - z_p\sqrt{\sigma_X^2 + \sigma_Y^2}$
- The (p,1-α) lower tolerance bound is:
  - *Exact solution if ratio of variances is known (Hall):*

$$\bar{X} - \bar{Y} - t_{n1+n2-2,1-\alpha}(z_p\sqrt{\nu})\frac{S_d}{\sqrt{\nu}}$$

$$\text{where } \nu = \frac{n_1(1+q)}{q + n_1/n_2}, \quad S_\text{d}^2 = \frac{\left(1+ {}^1\!/_q\right)\left[(n_1-1)S_X^2 + (n_2-1)qS_Y^2\right]}{n_1+n_2-2}, \text{ and } q = \frac{S_X^2}{S_Y^2}$$

  - *Approximate solution for unknown and arbitrary variances (Guo-Krishnamoorthy):*

$$\min\left(\bar{X} - \bar{Y} - t_{f_i,1-\alpha}(z_p\sqrt{\nu_i})\sqrt{\frac{S_X^2 + S_Y^2}{\nu_i}},\ i = 1,2\right)$$

$$\text{where } \nu_1 = \frac{n_1(1+q_1)}{q_1 + n_1/n_2}, \quad \nu_2 = \frac{n_2(1+q_2)}{q_2 + n_2/n_1}, \quad f_1 = \frac{(n_1-1)(q_1+1)^2}{q_1^2 + (n_1-1)/(n_2-1)}, \quad f_2 = \frac{(n_2-1)(q_2+1)^2}{q_2^2 + (n_2-1)/(n_1-1)}, \quad q_1 = \frac{S_X^2(n_2-3)}{S_Y^2(n_2-1)}, \quad q_2 = \frac{S_Y^2(n_1-3)}{S_X^2(n_1-1)}$$

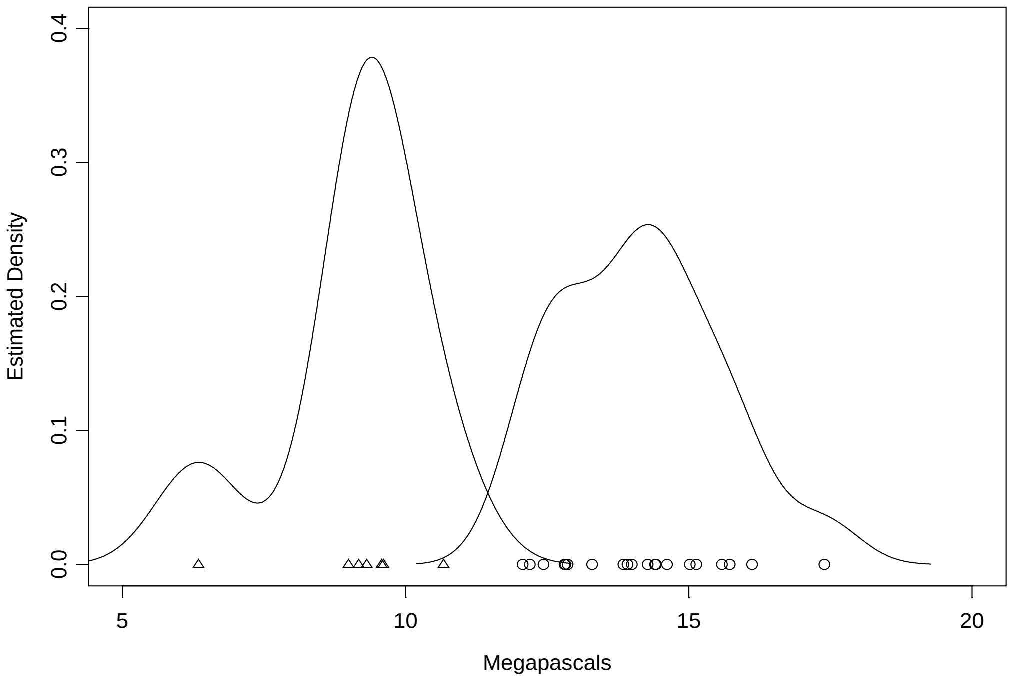
James.M.Womack@aero.org

CEV Parachute Assembly System (CPAS) Reliability

Load-Strength Reliability Problem (2)

- Simulated data: 20 strength, 7 load measurements:

Strength	16.1	17.4	14.6	12.8	14.0	15.0	14.3	12.9	15.7	14.4	15.6	13.3	13.8	13.9	12.2	12.4	12.8	14.4	15.1
Load	9.3	9.2	6.3	9.6	9.0	9.6	10.7												



- Approximate (p,.90) Lower Tolerance bounds

p	(p,.90) LTB
.99	-1.25
.98	-0.57
.97	-0.14
.9656	0.00
.96	0.17

- 90% confident that P(strength > load) > .9656

James.M.Womack@aero.org

AEROSPACE

Splus Code for Lower Tolerance Bound on X-Y

Ratio of Variances Known

```
# Computes (p,1-alpha) lower tolerance bound on X-Y for known ratio of variances
# inputs: xdata, ydata both vectors, p = percentile, alpha = 1 - confidence level, q1 = (sigma1/sigma2)^2

library(envstats)          # library for non-central t

Tolerance2distsKnownRatio <- function(p,alpha,q1,xdata,ydata)
{
        n1 <- length(xdata)
        n2 <- length(ydata)
        Sdsq <- (1+1/q1)*((n1-1)*var(xdata)+(n2-1)*q1*var(ydata))/(n1+n2-2)
        v1 <- n1*(1+q1)/(q1+n1/n2)
        LTB <- mean(xdata)-mean(ydata)-qt(1-alpha,n1+n2-2, ncp = qnorm(p)*sqrt(v1))*sqrt(Sdsq/v1)
        LTB
}
```


James.M.Womack@aero.org

Splus Code for Lower Tolerance Bound on X-Y

Unknown Arbitrary Variances

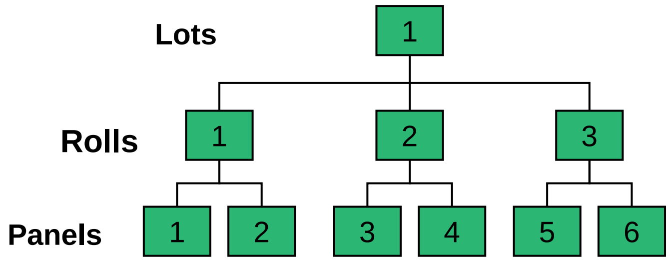
```
# Computes (p,1-alpha) lower tolerance bound on X-Y for unknown variances
# inputs: xdata, ydata both vectors, p = percentile, alpha = 1 - confidence level

library(envstats)          # library for non-central t

Tolerance2distsUnknownRatio <- function(p,alpha,xdata,ydata)
{
        n1 <- length(xdata)
        n2 <- length(ydata)
        S1sq <- var(xdata)
        S2sq <- var(ydata)
        q1 <- S1sq*(n2-3)/S2sq/(n2-1)
        q2 <- S2sq*(n1-3)/S1sq/(n1-1)
        v1 <- n1*(1+q1)/(q1+n1/n2)
        v2 <- n2*(1+q2)/(q2+n2/n1)
        f1 <- (n1-1)*(q1+1)^2/(q1^2+(n1-1)/(n2-1))
        f2 <- (n2-1)*(q2+1)^2/(q2^2+(n2-1)/(n1-1))
        LTB1 <- mean(xdata)-mean(ydata)-qt(1-alpha,f1, ncp = qnorm(p)*sqrt(v1))*sqrt((S1sq+S2sq)/v1)
        LTB2 <- mean(xdata)-mean(ydata)-qt(1-alpha,f2, ncp = qnorm(p)*sqrt(v2))*sqrt((S1sq+S2sq)/v2)
        LTB <- min(LTB1,LTB2)
}
```


James.M.Womack@aero.org

Structural Allowable for Silica Cloth Phenolic

Structured Data Example

- Silica Cloth Phenolic (SCP) is used as part of the RSRM and RSRM-V nozzles (TPS for flex bearing)
- Allowable stress (working stress or design allowable) for a material is the maximum stress at which one can be reasonably certain that failure will not occur
- Two one-sided tolerance limits are used
  - *A-basis: (.99,.95) lower tolerance bound on material strength*
  - *B-basis: (.90,.95) lower tolerance bound on material strength*
- Strength of composite materials may have several sources of variation
  - *Lot, Rolls, Panel*
  - *Temperature*
  - *Within panel variation*
  - *Other sources*



- Data with multiple sources of variability should be tested to determine if data is structured

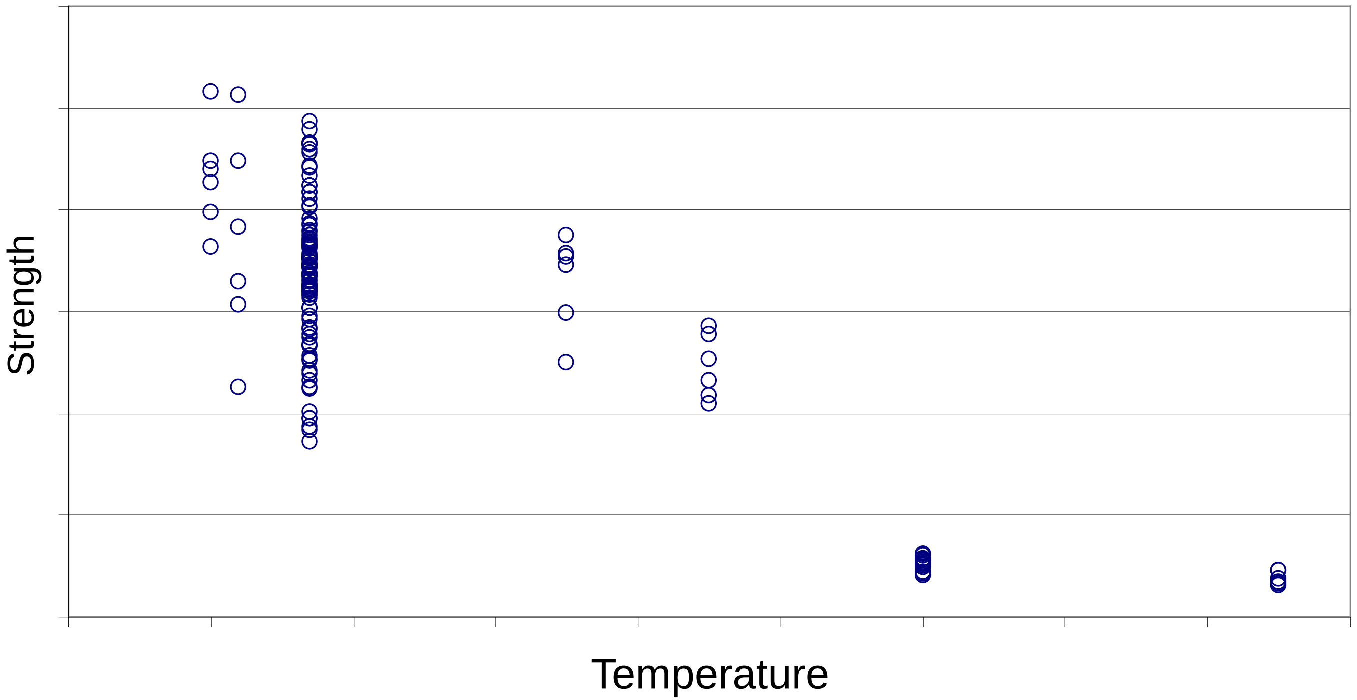
James.M.Womack@aero.org

AEROSPACE

Structural Allowable for Silica Cloth Phenolic

Structured Data Example (2)

- Silica Cloth Phenolic Panel Across-Ply Tensile Strength Data



- Objective: Determine lower tolerance bound at each temperature

James.M.Womack@aero.org

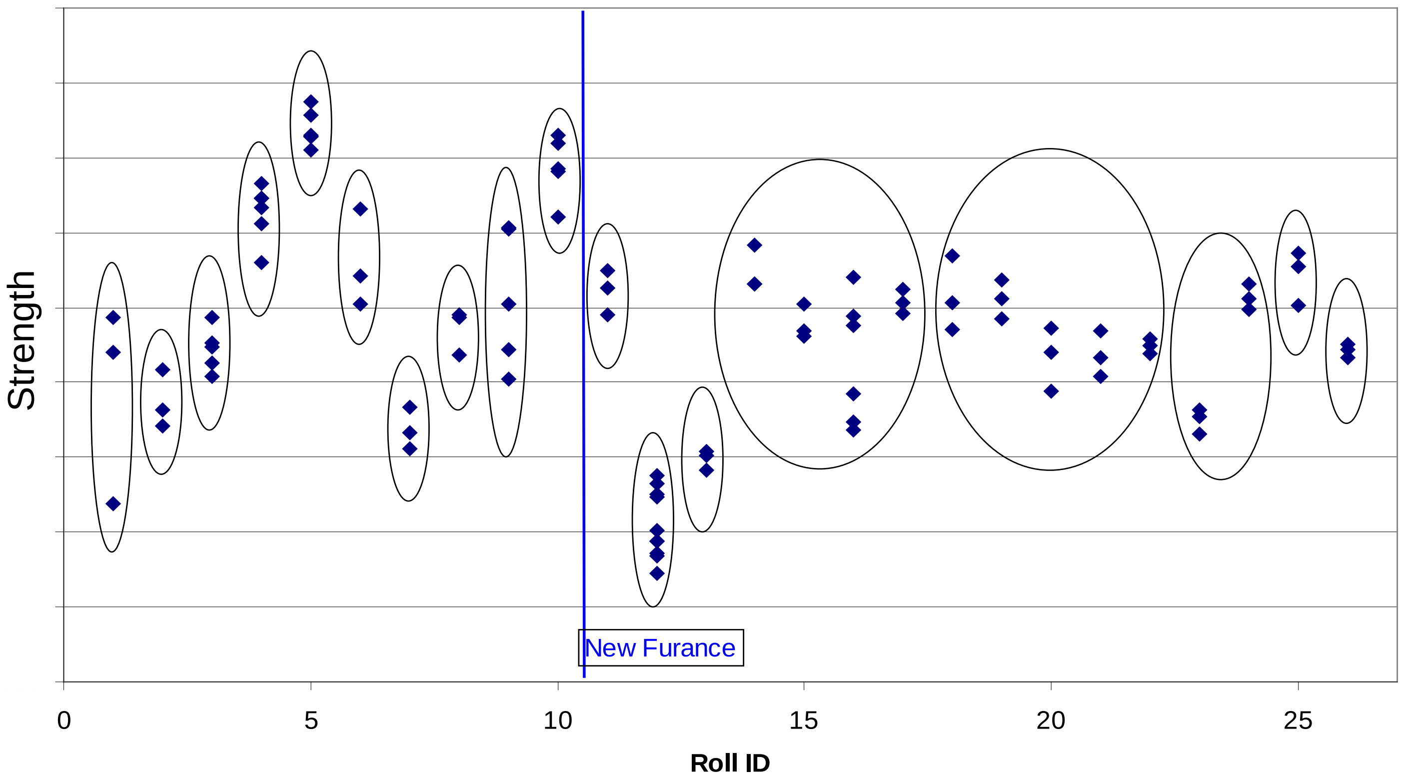
AEROSPACE

Structural Allowable for Silica Cloth Phenolic

Structured Data Example (3)

Silica Cloth Phenolic Panel Across-Ply Tensile Strength Data

Single Test Temperature



Anderson-Darling k-sample test proves the data is structured

James.M.Womack@aero.org

AEROSPACE

Estimating Data with 70 F Data Structure

Lot ID	Roll ID	Samples
3	3	5
4	4	6
5	5	6
6	6	3
7	7	3
8	8	3
9	9	5
10	10	5
11	11	3
12	12	10
13	13	3
14	14	3
	15	3
	16	6
	17	3
15	18	3
	19	3
	20	3
	21	3
	22	3
16	23	3
	24	3
17	25	3
18	26	3

- Randomly draw 16 lot means from $N(\mu,\sigma_{lot})$
- Randomly draw 24 roll means from $N(0, \sigma_{roll|lot})$
- Randomly draw 94 residual values from $N(0, \sigma_e)$
- Combine random data into lot, roll, sample data structure

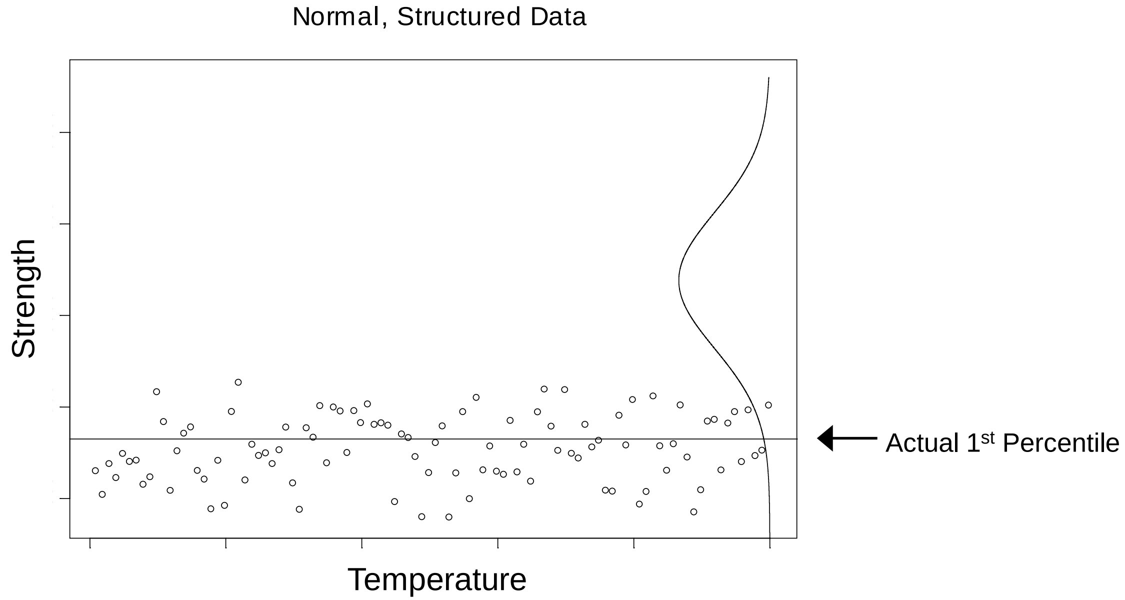
James.M.Womack@aero.org

AEROSPACE

Structural Allowable for Silica Cloth Phenolic

Structured Data Example (4)

- (.99,.95) normal tolerance bounds applied to structured data
  – *Using the estimated structure as in the SCP Tensile Strength data, 100 samples of size 94 are drawn and the tolerance bound is calculated*
  – *Simulated data includes variability between lot, within lots, within roll/panel*



- Normal tolerance bound is an overestimate approximately 35% of the time, not effective for structured data

James.M.Womack@aero.org

Mixed Effects Model For Silica Cloth Phenolic

Structured Data Example (5)

- Strength(T) = $\beta_0 + \beta_1 T + L + R + \varepsilon_k$ [Normal($\beta_0 + \beta_1 T$, $\sqrt{\sigma^2_{lot} + \sigma^2_{roll|lot} + \sigma^2_e}$)]
  - *T = temperature of SCP*
  - β_0, β_1 = slope and intercept parameters
  - *L = random lot effects, Normal(0, σ_{lot})*
  - *R = random roll effect within a lot, Normal(0, $\sigma_{roll|lot}$)*
  - *ε_k = Normal(0, σ_e)*
  - *Panel-to-panel variability ignored (only lot 14 roll 16 has more than 1 panel)*

- 1-Percentile of strength = $\beta_0 + \beta_1 T - 2.326 \sqrt{\sigma^2_{lot} + \sigma^2_{roll|lot} + \sigma^2_e}$

- (.99,.95) lower tolerance bound is a 95% lower confidence bound on 1-percentile
  - *Methods for computing tolerance bounds for this type of structured data are currently not available (unbalanced data with more than two variance terms)*
  - *Methods exist for unbalanced data with two variance terms and balanced data with more than two variance terms*

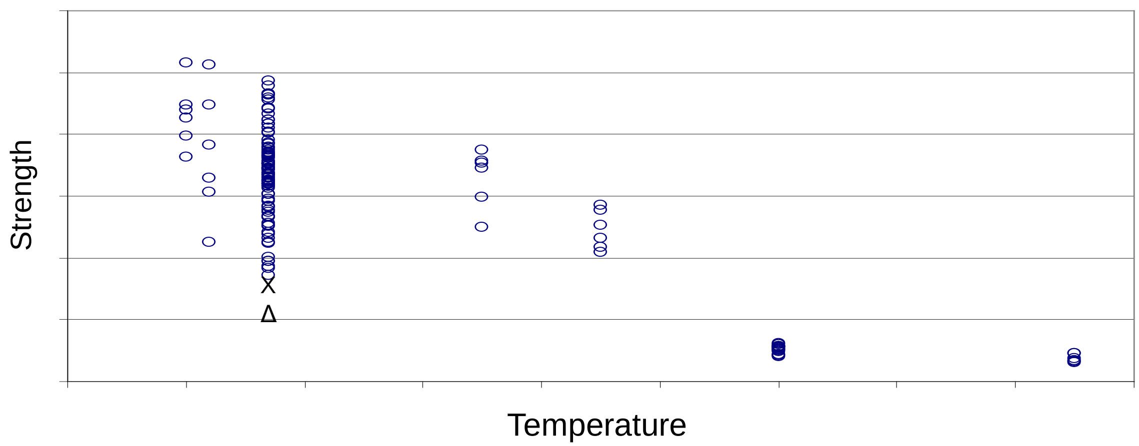
James.M.Womack@aero.org

AEROSPACE

Structural Allowable for Silica Cloth Phenolic

Structured Data Example (6)

- Silica Cloth Phenolic Panel Across-Ply Tensile Strength Data at 70 C
  - *Model as a mixed model with one random effect and unbalanced data*
    - Temperature is fixed effect
    - Combine lot and roll effect into single effect
    - Strength(T) = $\beta_0 + \beta_1 T + L + \varepsilon_k$ [Normal($\beta_0 + \beta_1 T, \sqrt{\sigma_{lot}^2 + \sigma_e^2}$)]
- Approximate methods exist for this case
  - *RECIPE ANOVA model used to compute (.99,.95) lower tolerance bound*
    - http://www.itl.nist.gov/div898/software/recipe/
  - *Splus code (based on Krishnamoorthy & Mathew)*



James.M.Womack@aero.org

Splus Code for Balanced Mixed Effects Models

```
# Computes Lower Tolerance Bounds for a Balanced Nested Design
# 1 Fixed Factor (Temperature) and 1 Random Factor (lot)
# Temperature has a levels and lot has b levels and n is the number of replicates
# Observed data is an array Y(i,j,el), i=1,2,...a, j=1,2,...,b, el=1,2,...,n
# p = the percentile of the tolerance bound
# alpha = 1 - confidence level of the tolerance bound

LTBmixedUnbalanced <- function(p,alpha,Y){
        a <- dim(Y)[1]
        b <- dim(Y)[2]
        n <- dim(Y)[3]
        f1 <- a*(b-1)
        f2 <- a*b*(n-1)
        d1 <- 1/(n*b)
        d2 <- (b-1)/(n*b)
        c1 <- 1/n
        c2 <- 1 - 1/n
        LTB <- rep(0,a)
        for (temp in 1:a){
                theta_hat <- sum(Y[temp,,])/(b*n)  # Step 1
                ss1 <- 0
                for (i in 1:a){
                        for(j in 1:b){
                                ss1 <- ss1 + ( sum(Y[i,j,])/n - sum(Y[i,,])/(b*n) )^2
                        }
                }
                ss2 <- 0
                for (i in 1:a){
                        for(j in 1:b){
                                for (el in 1:n){
                                        ss2 <- ss2 + ( Y[i,j,el] - sum(Y[i,j,])/n )^2
                                }
                        }
                }
                k <- 100000  # Step 2
                Z <- rnorm(k)  # Step 3
                U1 <- rchisq(k,f1)
                U2 <- rchisq(k,f2)
                G1 <- ss1/U1  # Step 4
                G2 <- ss2/U2
                Gtheta <- theta_hat - Z*sqrt(d1*G1 + d2*G2)
                G8 <- Gtheta - qnorm(p)*sqrt(c1*G1 + c2*G2)  # Step 5
                LTB[temp] <- quantile(G8,alpha)
        }
        LTB
}
```


James.M.Womack@aero.org

Summary

- Tolerance bounds are commonly used to verify requirements on space systems
  - *Determines limits of acceptability*
  - *Can also be used to determine required sample sizes*
  - *Includes uncertainty due to small sample sizes*

- Computing tolerance bounds
  - *First check distribution of data*
    - There are tolerance bounds for other types of distributions
  - *Check for structured data*
  - *Apply correct method*

- Research area: Tolerance bounds for general mixed effects or random effects models with unbalanced data

James.M.Womack@aero.org

References

- Hall, I. J. (1984), “Approximate one-sided tolerance limits for the difference of sum of two independent normal variates,” Journal of Quality Technology, 11, 13-19.
- Guo, H., Krishnamoorthy, K. (2004), “New Approximate Inferential Methods for the Reliability Parameter in a Stress-strength Model: The Normal Case,” Communications in Statistics - Theory and Methods, 33, 1715-1731.
- Krishnamoorthy, K., Mathew, T. (2009), Statistical Tolerance Regions: Theory, Applications, and Computation, Wiley.
- Vangel, M. G. (1995), “A Users Guide to RECIPE: A FORTRAN Program for Determining Regression Basis Values," http://www.itl.nist.gov/div898/software/recipe/recidoc.pdf

James.M.Womack@aero.org

Quality Engineering: A Journal Dedicated to Quality Improvement Methods and Applications

Dr. Connie M. Borror

Division of Mathematical and Natural Sciences

Arizona State University West;

Editor, Quality Engineering

ASU Arizona State University

Call for Papers

- **Special Issue on "Statistical Engineering" in *Quality Engineering***
- The objective of the issue is to invite contributions from the community of researchers, practitioners and academicians to provide examples, insight and research in the following areas:
  - Case studies demonstrating the development, process and results of the successful implementation of Statistical Engineering in a broad spectrum of different applications, including but not restricted to industry, manufacturing, service, financial, and healthcare
  - Implementation strategies to incorporate Statistical Engineering into the graduate and undergraduate Statistics program curricula
- ***Consider submitting a paper***
  - ***Handouts provided at this symposium on the how, when, where, what, and who!***

ASU Arizona State University

Purpose

- First issue: 1988
- Currently in Volume 23 (published quarterly)
- *Directed to* practitioners and researchers.
- *Devoted to* publication of original quality engineering solutions.
- Publish:
  - new methods ready for immediate application
  - novel uses of standard methods

ASU Arizona State University

Original Intentions…

Quality Engineering *is a magazine devoted to articles which tells persons dealing with Quality problems how others have addressed similar situations and what was done. The message should be "What the problem was, how we solved it, and what the results were."*

- Articles geared towards manufacturing-related issues.
- Journal was an outgrowth of the quality problems industry faced in the 1980s; collaboration between Marcel Dekker, Inc. and ASQC.

First Issue of QE

- “Do we need new machines? A p-chart and regression study” *Gerald B. Heyes*
- “New product introduction and quality program management” *James T. Zurn*
- “An application of fractional factorial experimental designs” *Mary B. Kilgo*
- “Management, measurement, and analysis of the supplier base” *Glenn Roth*
- “An approach for development of specifications for quality improvement” *Kailash C. Kapur*
- “New directions for reliability” *James R. King*
- “Nondestructive crimp verification” *James R. Simmonds*
- Variable gauge repeatability and reproducibility study using the analysis of variance method” *Pingfang Tsai*

ASU Arizona State University

Editors

- Frank Caplan, Founding Editor
  - Served for 17 years
- David Lyth
- James Simpson
- Geoff Vining
- Connie Borror
- Peter Parker (beginning January 2013)

ARIZONA STATE UNIVERSITY

QE Sections

- Original Articles
- *Quality Quandries*
- *Statistical Standards*
- *Technical Advice (new)*
- *Reliability Section (new)*
  - *We need more articles here*
  - *The Reliability Division of ASQ gives a $1000 award for the best reliability paper in QE each year.*

1988-2005

Certified quality engineer–Body of knowledge categories

No.	Category description	%
I.	Management and leadership in quality engineering	14.8
II.	Quality systems development, implementation, and verification	1.7
III.	Planning, controlling, and assuring product and process quality	17.6
IV.	Reliability and risk management	4.6
V.	Problem solving and quality improvement	4.6
VI.	Quantitative methods	56.7

Booker, B. and Lyth, D. (2006). Editorial: "*Quality Engineering* from 1988 through 2005; Lessons from the Past and Trends for the Future". Quality Engineering pp. 1-4.

1988-2005

Quantitative methods		
VI.	Sub-category description	%
a.	Concepts of probability and statistics	1.8
b.	Collecting and summarizing data	4.6
c.	Properties and applications of probability distributions	4.1
d.	Statistical decision-making	3.2
e.	Measuring and modeling relationships between variables	12.0
f.	Designing experiments	27.9
g.	Statistical process control (SPC)	34.2
h.	Analyzing process capability	12.4
i.	Concepts of probability and statistics	1.8

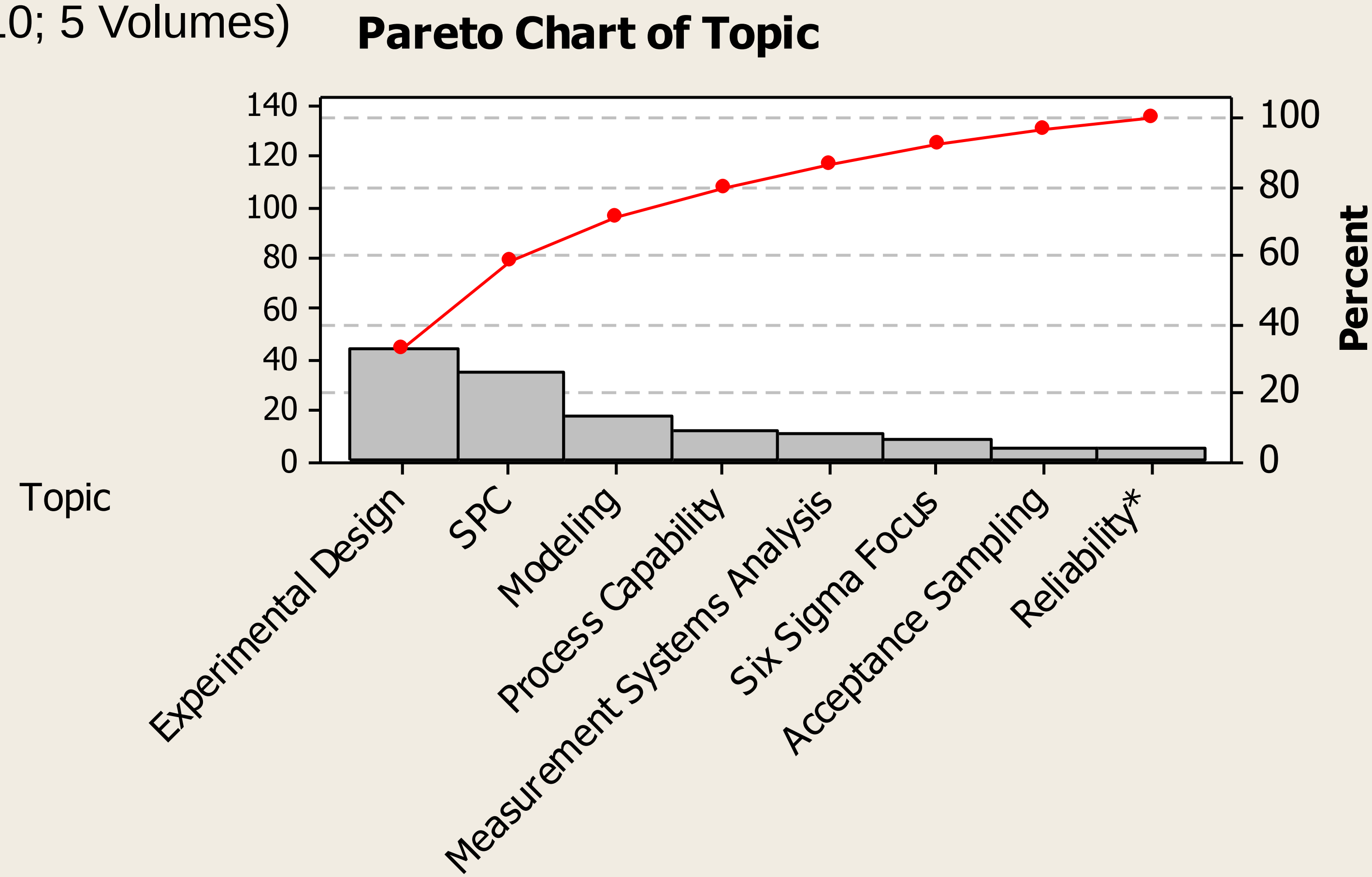
Booker, B. and Lyth, D. (2006). Editorial: “*Quality Engineering* from 1988 through 2005; Lessons from the Past and Trends for the Future”. Quality Engineering pp. 1-4.

Topics

(2006-2010; 5 Volumes)

N = 152

$R^+ = 3$



Number of Articles	44	35	17	11	10	8	5	5	= 135
Percent	32.6	25.9	12.6	8.1	7.4	5.9	3.7	3.7	
Cum %	32.6	58.5	71.1	79.3	86.7	92.6	96.3	100.0	

Articles

- Original research
- Novel applications
- Tutorials
- Review papers
- Discussion papers
- Historical
- “Conversation series”

Arizona State University

What do we need?

- Collaboration among different fields and expertise

In academia:

- When times are good: (these are turtles):

We're happy collaborators

Engineering

Math/statistics

Business

Chemistry/Biology

….

ARIZONA STATE UNIVERSITY

http://freeturtle.net/freeturtle

What do we need?

- Collaboration among different fields and expertise is necessary

In academia:

- When times are bad, we have to be careful: (these are turtles also):

Engineering

Math/statistics

Business

Chemistry/Biology

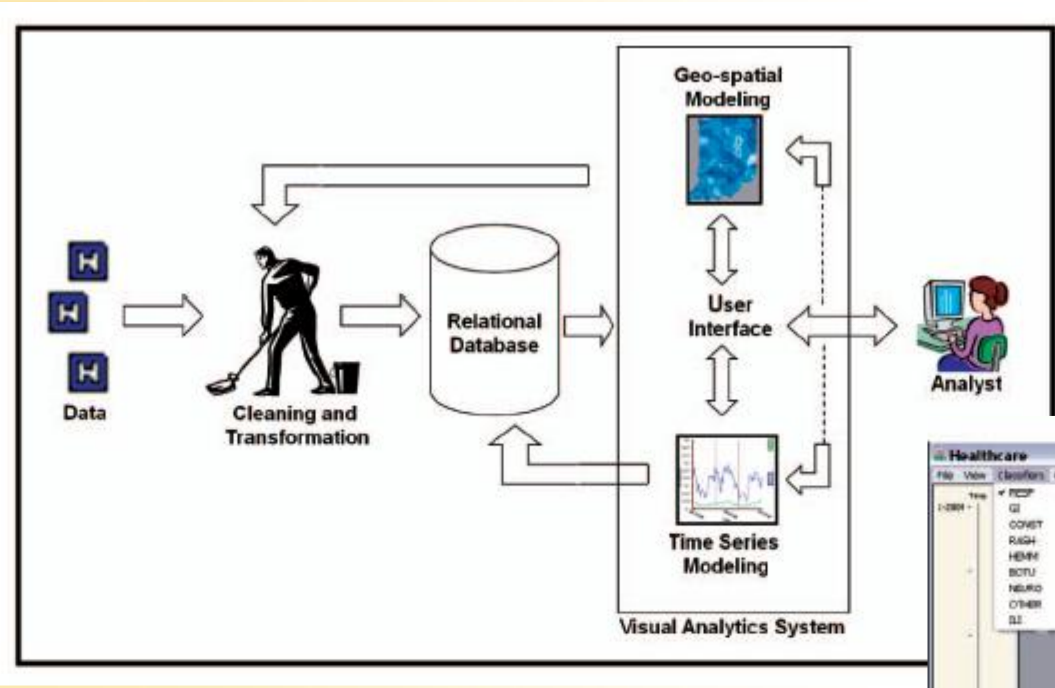
….

We “turtle up”

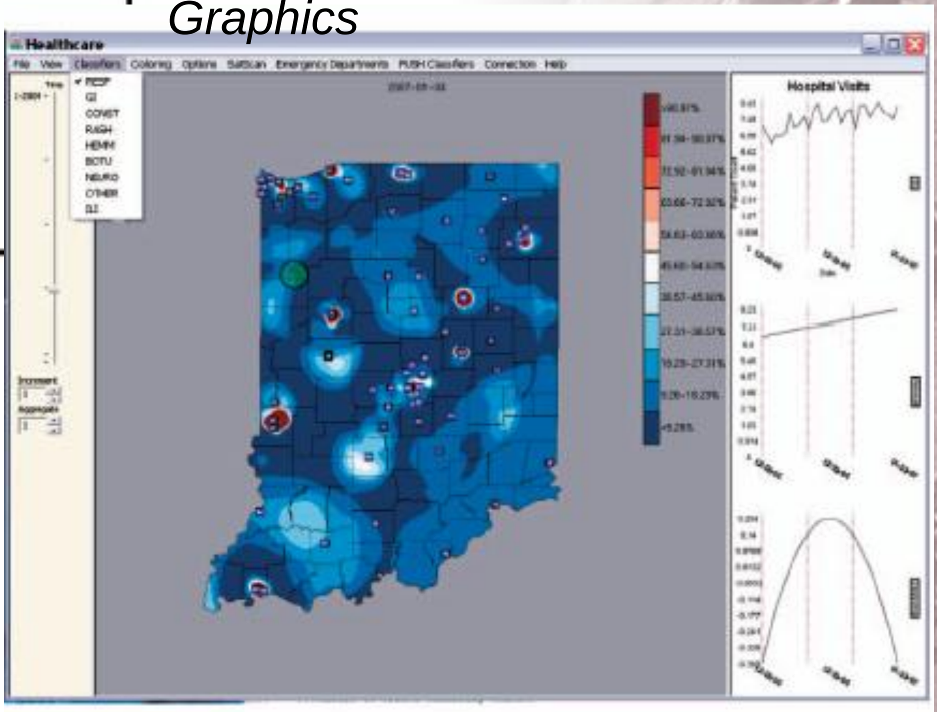
http://freeturtle.net/freeturtle

Collaboration

- Best methods + Software = Useful + Used



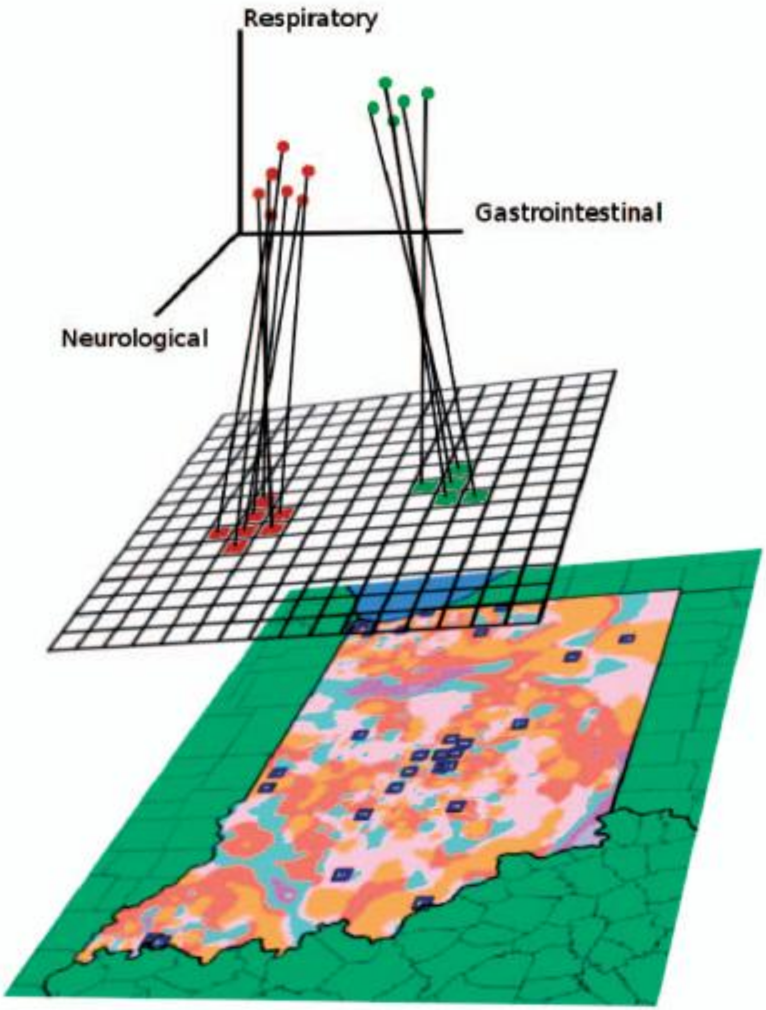
Maciejewski, Rudolph, Hafen, Abusalah, Yakout, Ouzzani, Cleveland, Grannis, and Ebert, (2010). “Visual Analytics Approach to Understanding Spatiotemporal Hotspots”. *IEEE Transactions on Visualization and Computer Graphics*



Collaboration

- Best methods + Software = Useful + Used

Maciejewski, Rudolph, Hafen, Larew, Mitchell, Cleveland, and Ebert, (2010). “Forecasting Hotspots – A Predictive Analytics Approach”*. IEEE Transactions on Visualization and Computer Graphics*



ARIZONA STATE UNIVERSITY

What do we want in QE?

- Every so often, we need
  - Bibliographies
  - Basic review articles
- Immediate need (want?)
  - Basic review of reliability techniques and methods
  - Providing references for “where to start”
- Software reviews/guidance
- What’s it called in that field?
  - Key to guiding students as well as practitioners
  - Requires collaboration among experts in these various fields

ARIZONA STATE UNIVERSITY

Quality Engineering

- Fills a much needed niche
- Outlet for innovative collaborations
- By its original intent, QE embodies “statistical engineering”
- Teaching tool

Quality Engineering

- No matter what the application or field of interest, *QE* articles demonstrate:

  “*What the problem was, how we solved it, and what the results were.*”

- I would like to dedicate this presentation in memory of a beloved leader in this field and unfailing supporter of *QE*, Dr. Soren Bisgaard.

Arizona State University

Experience with Designed Experiments in Aerospace Ground Test: Successes and Challenges

Dr. Drew Landman
Professor
Department of Mechanical and Aerospace Engineering
Old Dominion University

Introduction

- The history of the use of DOE in aerospace ground testing appears to be rather short
  - Examples
    - NASA LaRC began use in the late 90's
    - USAF AEDC only recently began training in DOE
- Personal experience
  - Began in summer 2001 by setting up wind tunnel experiments for DOE classwork
  - Experience now with DOE use in aerodynamic testing of aircraft, automobiles, trucks as well as instrument calib.
  - Currently teaching DOE/RSM at graduate level to ODU students and conducting industry/gov. training
  - This talk will highlight some of the author's experiences with DOE based problems

OLD DOMINION UNIVERSITY

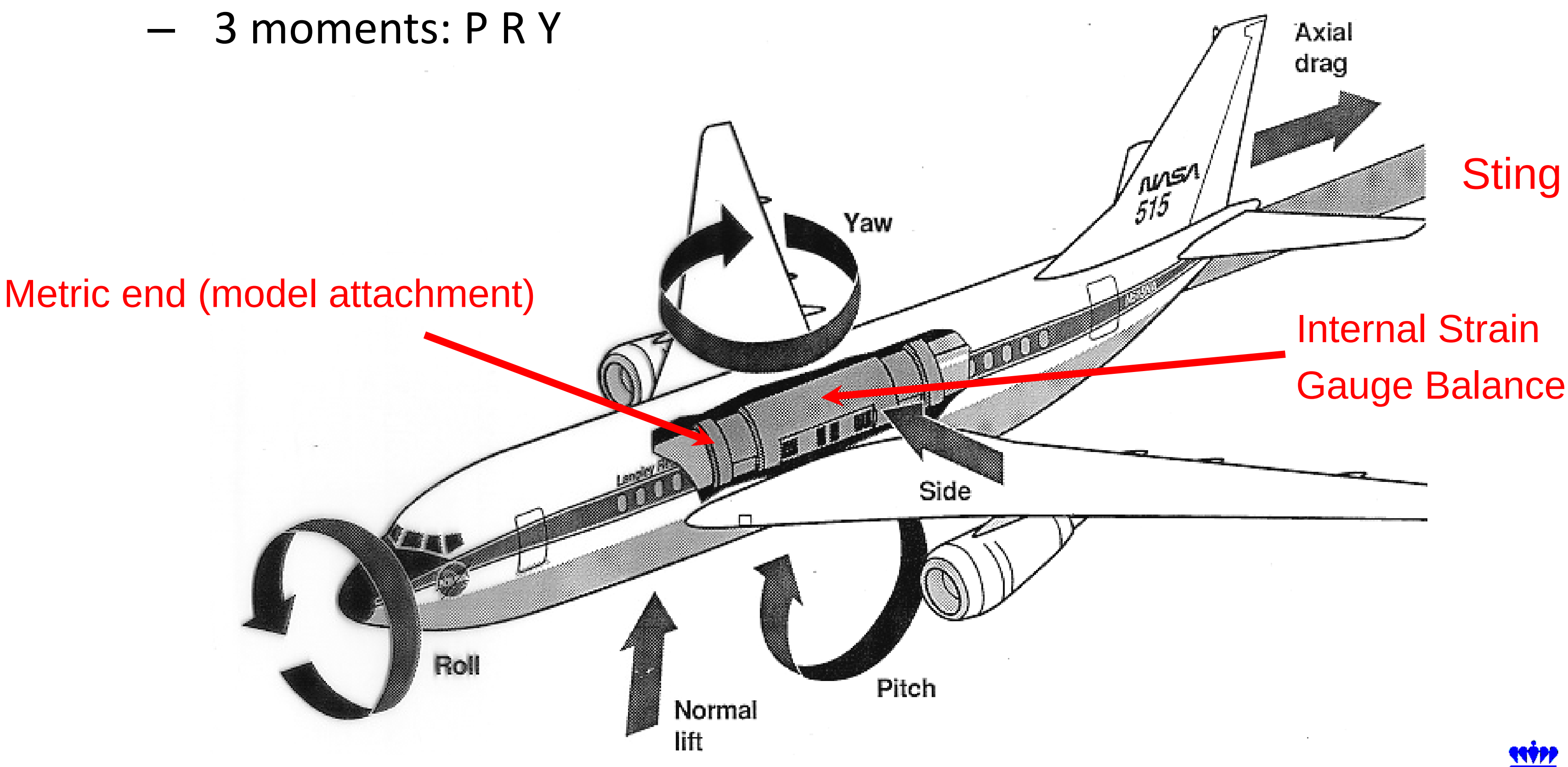
Internal Balance Calibration

- Force and moment measurements are fundamental to wind tunnel testing (two LFST examples below)
- Recently more emphasis on determining uncertainty
- Calibration is required to develop a mathematical model used to predict aerodynamic loading
- Methodology involves an experimental approach executed with a mechanical load system
  - Both manual and automated

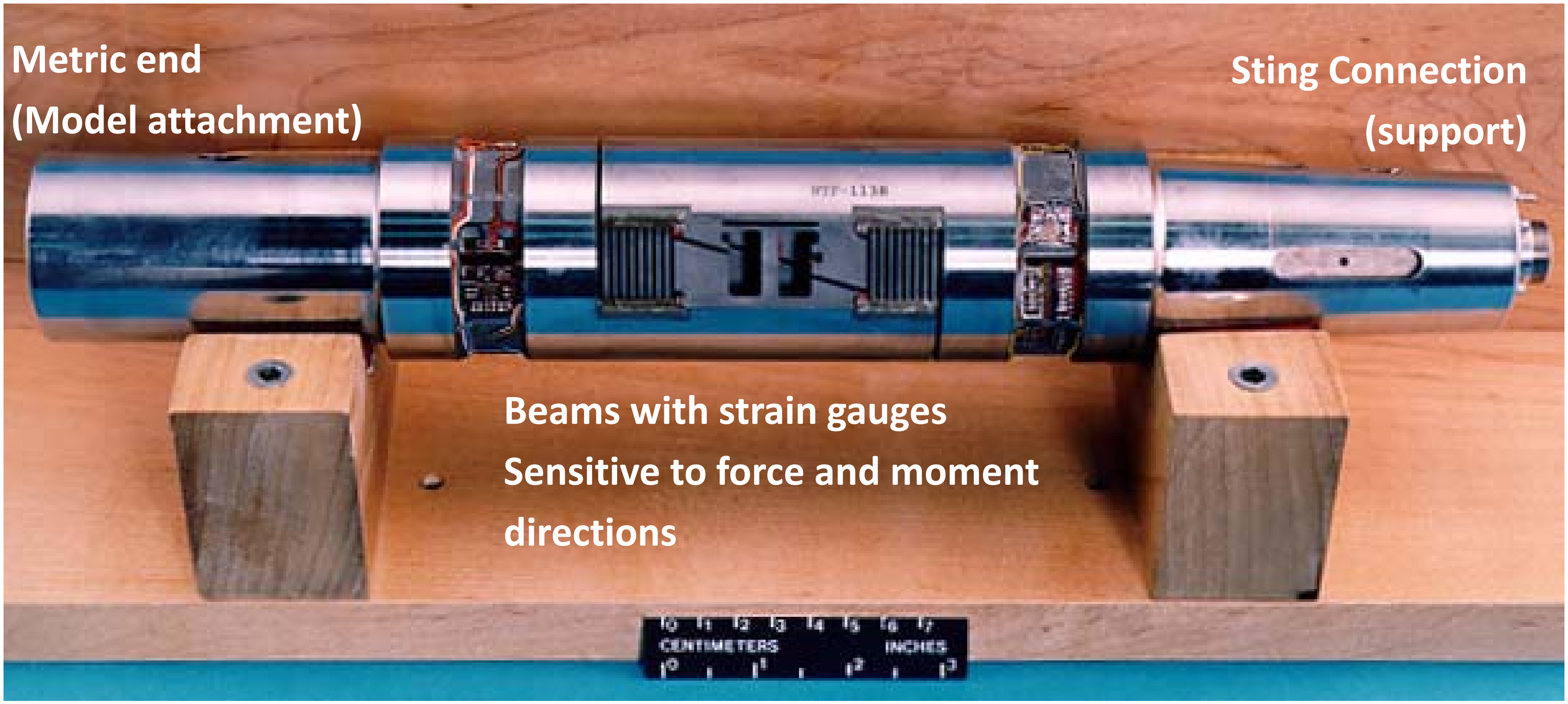
OLD DOMINION UNIVERSITY

Internal Balance Measurements

- Internal balance located in aircraft model
- 6 degrees of freedom, Measures:
  - 3 forces: N A S
  - 3 moments: P R Y



Internal Strain Gauge Balance



Input: Aerodynamic Forces (3) and Moments (3)

Output: 6 Voltages from excited strain gage bridges

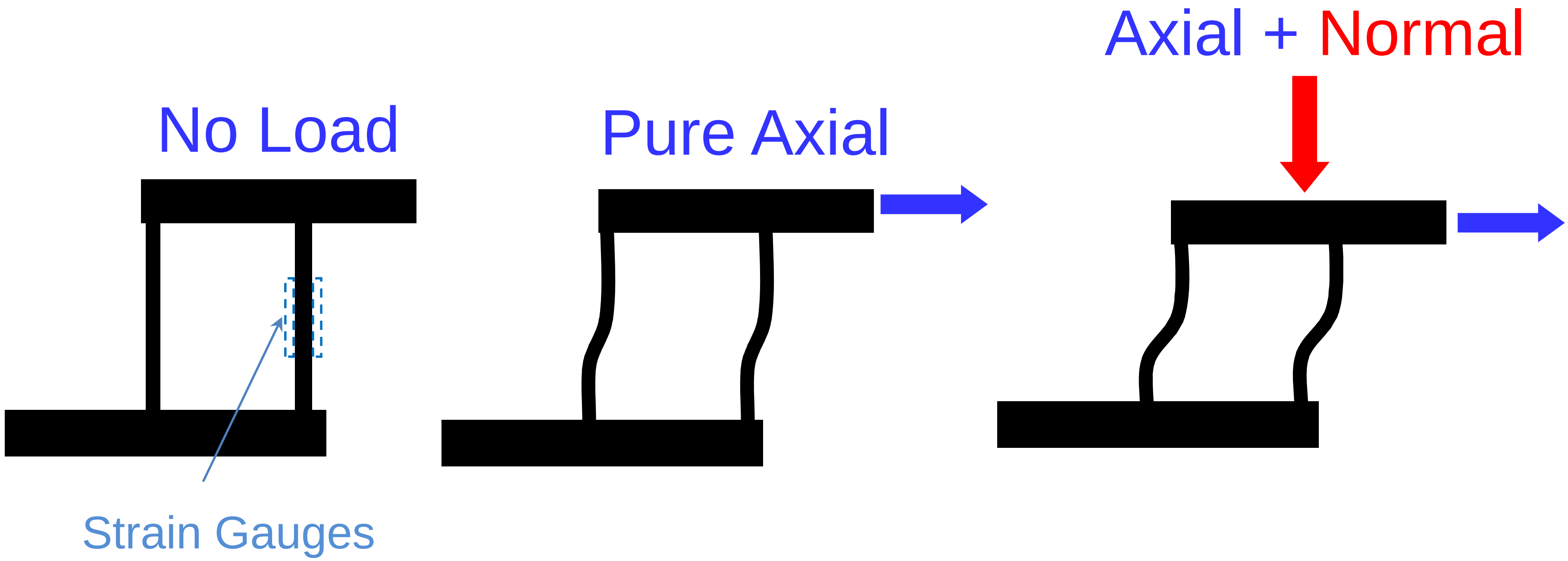
Balance Calibration

- Calibration method
  - Apply known combinations of loadings and collect voltage responses from the balance
  - Build an empirical math model
- No pure loads – interactions will occur
  - Structure design to best characterize interactions
- Historically has been very labor intensive and time consuming

OLD DOMINION UNIVERSITY

Interactions

- Look at a pair of beams for measuring load in Axial direction but subject to simultaneous loading in Normal direction



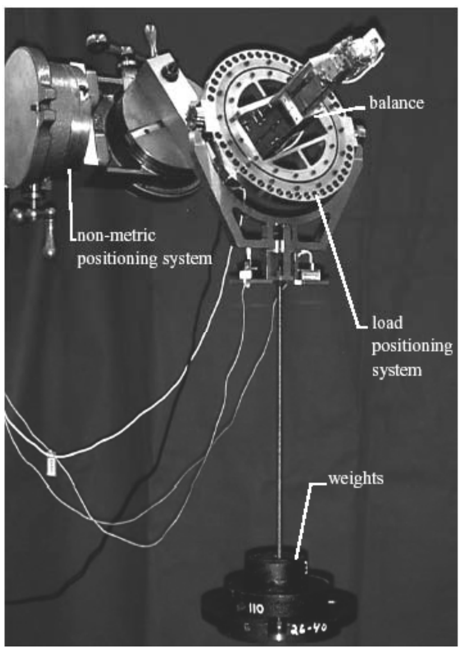
- 27 term math model for each component, example of Axial force

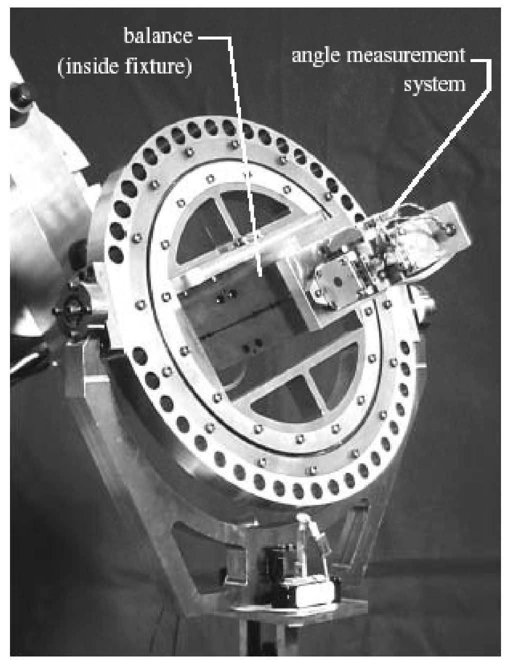
$$Axial(volts) = \beta_A A + \beta_N N + \beta_S S + \beta_P P + \beta_R R + \beta_Y Y$$
$$+ \beta_{AN} AN + \beta_{AS} AS \ldots + \beta_{AA} A^2 + \beta_{NN} N^2 + \ldots$$

OLD DOMINION UNIVERSITY

Automated Balance Calibration Example

- NASA LaRC Single Vector System (SVS)
- DOE-based calibration – Central Composite Design
- Automated calibration machine
- Typically 24 hour calibration duration
- Currently limited to balance capacities <u>less than 3000 lbs</u>





Manual Calibration Methods

- Traditionally done with dead weight loading in fixtures with cables and levers to apply loads and moments
- Time consuming – traditional cal takes 3 weeks or more
- Required for high capacity balances

Old Dominion University

High-Capacity Balance Calibration Process

- Traditional Experimental Approach
  - Extensive load matrix changing one axis at a time – 729 points
  - Data used to build second order empirical models
  - Proven, accepted method in use for 50 years

- Mechanical Calibration System
  - Reliable, dependable, accurate
  - Heavy loads, slow process
    – 3 to 4 weeks
  - Consider a DOE approach

OLD DOMINION UNIVERSITY

Test Matrix Issues

- Which combinations of forces and moments?
  - Ideally, simultaneous loading in all forces and moments for each run (and changing after each) is desired
    - Limitation of manual calibration stand – level of difficulty increases from one to six simultaneous loads
  - The original 729-point uses 5 levels in a pyramid scheme
  - SVS uses 5 levels in modified CCD
  - Quadratic models require only 3 levels
- What about the test sequence?
  - Randomized test sequence is ideal to minimize the effects of unwanted variability on model estimation
  - Increases overall test time due to assembly configuration changeovers

Chosen Design Approach

- Based on modified Box-Behnken design
- 3 levels hi,mid,low
- Primarily two factors at a time
  - Mechanical constraints force several > 2
- Total of 65 loadings versus traditional 729
- Chose off-center zero run
  - Could be omitted and left as zero loading for further efficiency

Normal	Axial	Pitch	Roll	Yaw	Side	Runs
±1	±1	0	0	0	0	4
±1	0	±1	0	0	0	4
±1	0	0	±1	0	0	4
±1	0	0	0	±1	±1	4
±1	0	0	0	0	±1	4
0.13	±1	±1	0	0	0	4
0.06	±1	0	±1	0	0	4
0	±1	0	0	±1	0.13	4
0	±1	0	0	0	±1	4
0.19	0	±1	±1	0	0	4
0.13	0	±1	0	±1	±1	4
0	0	±1	0	0	±1	4
0	0	0	±1	±1	0.25	4
0	0	0	±1	0	±1	4
0	0	0	0	±1	±1	4
0.10	0	0.10	0.10	0.10	0.10	5

Trial Calibration Summary

- Trial modified BBD took 90 hours (low capacity balance used)
- Traditional 729 OFAT cal on same balance takes 120 hours
  - For a high capacity balance it takes approximately 200-240 hours
  - Time savings should be much more significant with the heavy capacity balance and new method, estimated 40-50% time savings

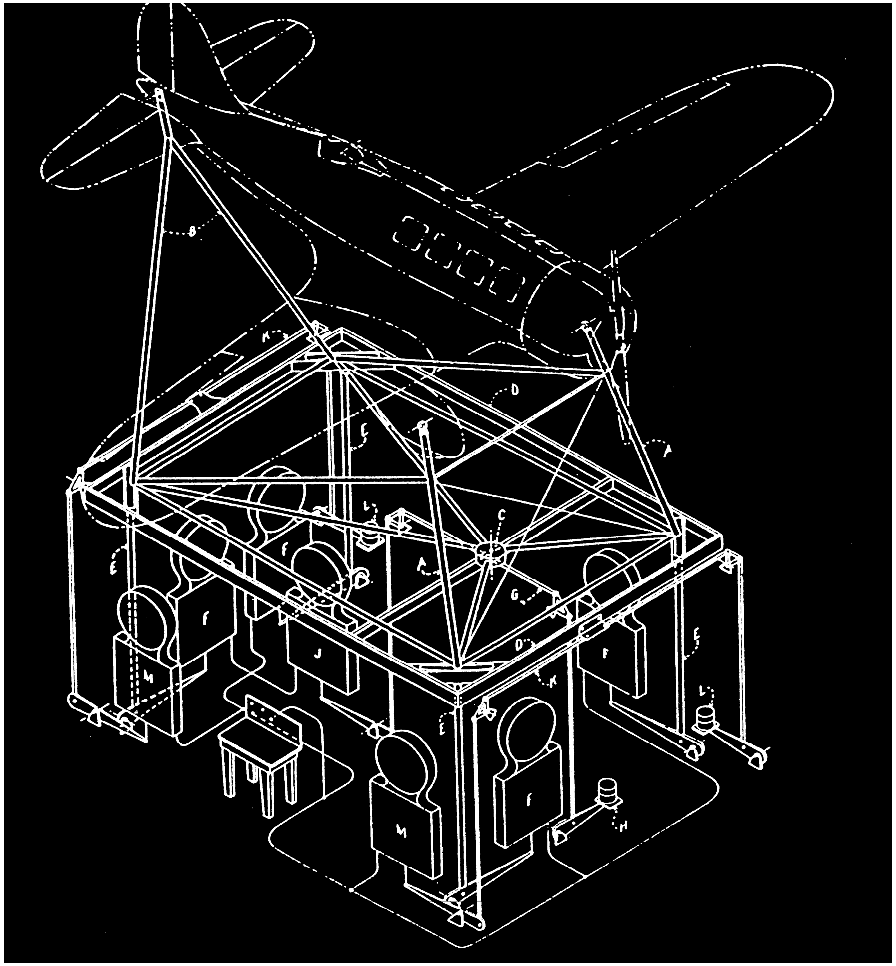
- In addition, new design has
  - Higher statistical power than traditional design in ability to estimate model terms
  - Robust estimates of uncertainty

- Challenges - Culture
  - Convincing technicians and managers that:
    - Randomization is required
    - 3 levels are enough to define the desired quadratic model

Old Dominion University

External Balance Calibration

- External balance is *external* to aircraft/ model
- Linkages between model and platform/ platform and scales
- Example from NASA LaRC Full-Scale Tunnel (later called LFST)
- Original engineering drawing form 1930's



OLD DOMINION UNIVERSITY

Inspiration

- Fall of 2005
  - Minor flooding required rebuilding of FST external balance
  - Boeing Phantom Works hires ODU to test X-48B in spring of 2006
- Fresh calibration required
  - Small differences in link lengths and relative angles as result of refurbishment

Old Dominion University

Calibration Approach

- Differs from internal balance
  - Calibration must be done in the facility
    - Desire to minimize downtime of facility
    - No standard loading apparatus (compared to cal. stands)
  - The traditional calibration model is 1st order without interactions
  - Easier to perform multi-component loading (min deflection)
- Common to both approaches, require
  - Sufficient statistical power for model term estimation
  - Robust estimates of uncertainty
  - Minimum number of loadings
  - Provision for interactions
  - Test for model adequacy

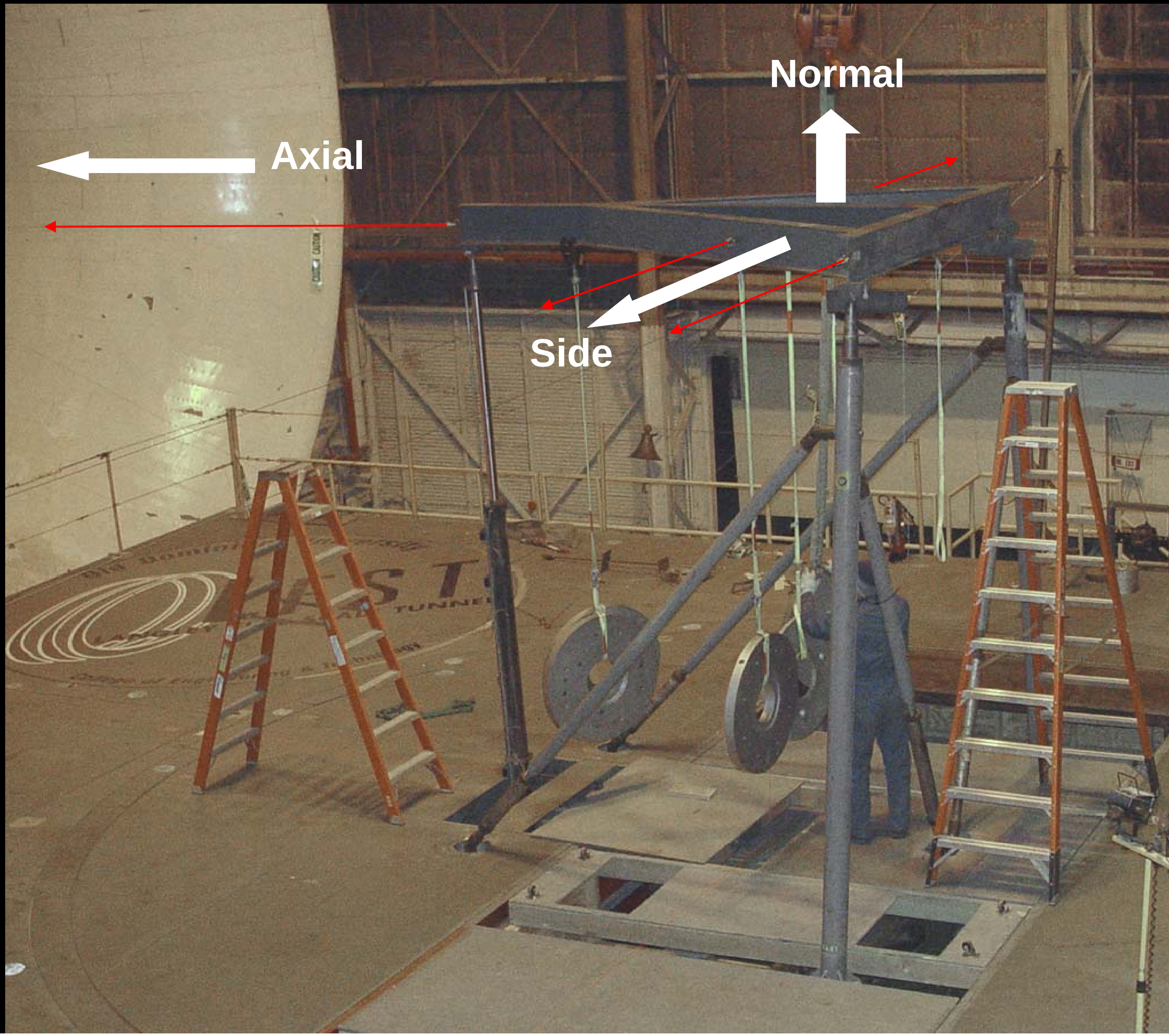
OLD DOMINION UNIVERSITY

Calibration Load Frame

- A rigid frame is positioned on the wind tunnel aircraft support struts to allow multi-component loading
- Weights are hung below to generate pitch & roll moments, normal force
- Cables and pulleys are used to generate side and axial force, yaw moment
- Load frame and cables must be level/orthogonal
  - Usually only have to do this once as compared to internal balance cal where it has to be adjusted for every run

Old Dominion University

Load Frame



Design Choice

- A 2_{VI}^{6-1} two-level fractional factorial design
  - Run-efficient 32 + centers and validation points
  - Provides minimum desired model order (from past cal)

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j}\sum \beta_{ij} x_i x_j + \varepsilon$$

- Fully randomize the test matrix
  - Protect against lurking vars. e.g. temp
- Load all components simultaneously
  - Run efficient, identifies interactions
- Modifications were required to the ideal design points due to finite load combination limits

OLD DOMINION UNIVERSITY

Test Matrix Development

- Sufficient statistical power
  - drives number of runs chosen
- Weigh-beam balance design means relative loadings
  - 0 to 1000 lbs or -500 to +500 same results
- Load ranges chosen
  - Predicted loads on X-48B
  - Availability of weights
  - Distribution of multi-component loadings

Old Dominion University

Ideal vs. Achievable

Std Order	Run Order	Ideal Half Fraction: Normal	Axial	Side	Pitch	Roll	Yaw	Achievable Factor Levels: Normal	Axial	Side	Pitch	Roll	Yaw
33	1	-500.0	75.0	25.0	1440.0	721.0	177.7	-500.4	80.0	30.0	2312.9	721.0	123.3
10	2	0.0	0.0	0.0	3400.0	1.5	-4.6	0.0	0.0	0.0	3234.5	1.5	0.0
36	3	-500.0	75.0	25.0	1440.0	721.0	177.7	-500.4	80.0	30.0	2312.9	721.0	123.3
15	4	-1000.0	150.0	50.0	3400.0	1.5	360.0	-1000.9	150.0	50.0	3496.2	443.8	358.0
20	5	0.0	150.0	0.0	-513.3	1229.0	360.0	0.0	150.0	0.0	29.2	1.5	302.6
35	6	-500.0	75.0	25.0	1440.0	721.0	177.7	-500.4	80.0	30.0	2312.9	721.0	123.3
30	7	0.0	0.0	50.0	3400.0	1229.0	-4.6	0.0	0.0	50.0	3234.5	1229.0	55.5
4	8	0.0	150.0	0.0	-513.3	1.5	-4.6	0.0	150.0	0.0	29.2	1.5	-4.6
17	9	-1000.0	0.0	0.0	-513.3	1229.0	360.0	-1000.9	0.0	0.0	-227.0	1086.9	307.2
25	10	-1000.0	0.0	0.0	3400.0	1229.0	-4.6	-1000.9	0.0	0.0	2082.5	1088.4	0.0
34	11	-500.0	75.0	25.0	1440.0	721.0	177.7	-500.4	80.0	30.0	2312.9	721.0	123.3
9	12	-1000.0	0.0	0.0	3400.0	1.5	360.0	-1000.9	0.0	0.0	2857.5	-31.0	307.2
22	13	0.0	0.0	50.0	-513.3	1229.0	360.0	0.0	0.0	50.0	0.0	1227.5	362.7
16	14	0.0	150.0	50.0	3400.0	1.5	-4.6	0.0	150.0	50.0	3860.5	102.9	50.8
23	15	-1000.0	150.0	50.0	-513.3	1229.0	360.0	-1000.9	150.0	50.0	-513.3	1193.9	358.0
31	16	-1000.0	150.0	50.0	3400.0	1229.0	-4.6	-1000.9	150.0	50.0	3455.3	1196.4	50.8
29	17	-1000.0	0.0	50.0	3400.0	1229.0	360.0	-1000.9	0.0	50.0	2870.2	1197.9	362.7
5	18	-1000.0	0.0	50.0	-513.3	1.5	360.0	-1000.9	0.0	50.0	-227.0	77.0	362.7
7	19	-1000.0	150.0	50.0	-513.3	1.5	-4.6	-1000.9	150.0	50.0	-513.3	76.0	50.8
26	20	0.0	0.0	0.0	3400.0	1229.0	360.0	0.0	0.0	0.0	3234.5	1119.5	307.2
28	21	0.0	150.0	0.0	3400.0	1229.0	-4.6	0.0	150.0	0.0	3072.8	1119.5	-4.6
21	22	-1000.0	0.0	50.0	-513.3	1229.0	-4.6	-1000.9	0.0	50.0	-227.0	1194.9	55.5
3	23	-1000.0	150.0	0.0	-513.3	1.5	360.0	-1000.9	150.0	0.0	-513.3	-31.0	302.6
6	24	0.0	0.0	50.0	-513.3	1.5	-4.6	0.0	0.0	50.0	0.0	109.5	55.5
14	25	0.0	0.0	50.0	3400.0	1.5	360.0	0.0	0.0	50.0	3234.5	109.5	362.7
12	26	0.0	150.0	0.0	3400.0	1.5	360.0	0.0	150.0	0.0	3682.4	0.0	302.6
24	27	0.0	150.0	50.0	-513.3	1229.0	-4.6	0.0	150.0	50.0	29.2	1226.4	50.8
1	28	-1000.0	0.0	0.0	-513.3	1.5	-4.6	-1000.9	0.0	0.0	-227.0	-32.5	0.0
2	29	0.0	0.0	0.0	-513.3	1.5	360.0	0.0	0.0	0.0	0.0	0.0	307.2
11	30	-1000.0	150.0	0.0	3400.0	1.5	-4.6	-1000.9	150.0	0.0	3455.3	-32.5	-4.6
32	31	0.0	150.0	50.0	3400.0	1229.0	360.0	0.0	150.0	50.0	3072.8	1229.0	358.0
13	32	-1000.0	0.0	50.0	3400.0	1.5	-4.6	-1000.9	0.0	50.0	3467.0	78.5	55.5
19	33	-1000.0	150.0	0.0	-513.3	1229.0	-4.6	-1000.9	150.0	0.0	-513.3	1088.4	-4.6
18	34	0.0	0.0	0.0	-513.3	1229.0	-4.6	0.0	0.0	0.0	0.0	1119.5	0.0
27	35	-1000.0	150.0	0.0	3400.0	1229.0	360.0	-1000.9	150.0	0.0	3455.3	1086.9	302.6
8	36	0.0	150.0	50.0	-513.3	1.5	360.0	0.0	150.0	50.0	29.2	102.9	358.0

Minor Effects on Chosen Model

- Power is only slightly effected at the desired 2 std dev level
  - ME's 99.9 vs. low 96.1
  - 2FI's 99.9 vs. one low of 87
- VIF evaluates multicolinearity
  - R.O.T desire <10
  - 1.18 is the worst case here

		Power at α = 5%	
Term	**VIF**	**1 Std. Dev.**	**2 Std. Dev.**
A	1.05	70.9 %	99.9 %
B	1.08	69.7 %	99.9 %
C	1.07	70.5 %	99.9 %
D	1.15	46.3 %	96.1 %
E	1.11	56.6 %	98.9 %
F	1.09	55.6 %	98.7 %
AB	1.08	69.9 %	99.9 %
AC	1.09	69.3 %	99.8 %
AD	1.11	47.0 %	96.4 %
AE	1.10	56.8 %	98.9 %
AF	1.08	55.5 %	98.7 %
BC	1.07	70.1 %	99.9 %
BD	1.11	46.5 %	96.2 %
BE	1.14	55.3 %	98.6 %
BF	1.10	54.5 %	98.5 %
CD	1.07	48.1 %	96.8 %
CE	1.06	57.8 %	99.0 %
CF	1.05	55.8 %	98.7 %
DE	1.18	35.1 %	88.1 %
DF	1.14	34.0 %	87.0 %
EF	1.14	42.6 %	94.2 %

Old Dominion University

Calibration Summary

- Successes
  - 32 factorial runs + 4 centers + 5 validation points = 41 total loadings
    - Passed test for lack of fit
    - If a higher order model is required, easy to augment to CCD
    - Future calibrations or check cals could use 2_{IV}^{6-2} as only a few 2FI's found
  - Minor changes in set points from standard design had negligible impact on model quality
  - 13 hours vs. 40 hours of previous OFAT calibration
  - Robust validation of first order + 2 FI model and uncertainty level estimation
  - A successful test of the X-48B and later X-48C followed the calibration

Old Dominion University

Calibration Summary

- Challenges – Mostly Cultural
  - Simultaneous multi-component loading requires that more weights be available versus previous OFAT design
  - Engineers have traditionally felt that a load sequence involving 5 or more levels is required
    - to characterize an essentially first order system ?
  - Technicians have to be taught to accept the fully randomized test sequence
    - Always a struggle as they think they are "doing it the hard way"

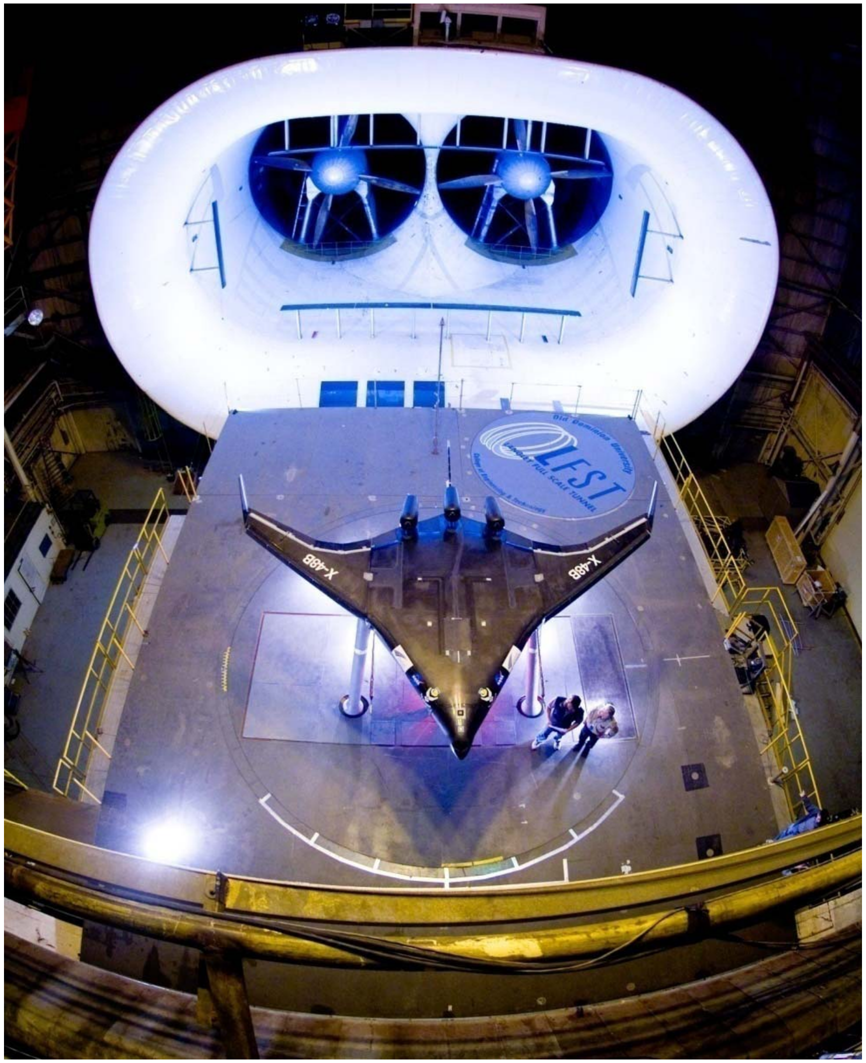
Background: Blended Wing Body at the LFST

Year long test program 2005-2006

- NASA Static model testing
- NASA Free flight model testing
- Boeing X-48 B flight model

3 Month Program in 2009

- Boeing X-48 C flight model



OLD DOMINION UNIVERSITY

Typical Low-Speed Wind Tunnel Test Objectives

- Aerodynamic Characterization
  - A math model describing each response in terms of all the factors i.e. $C_D = C_D(\alpha, \beta, \delta_1, \delta_2, \ldots)$
    - Stability and Control analysis
    - Computer flight simulations
  - Limits to flight (eg. stall)
- Capture the true uncertainty in testing
  - The sum of error due to:
    - Setting the control surfaces
    - Setting attitude
    - Force balance variances
    - Dynamic pressure measurement
    - Environmental effects

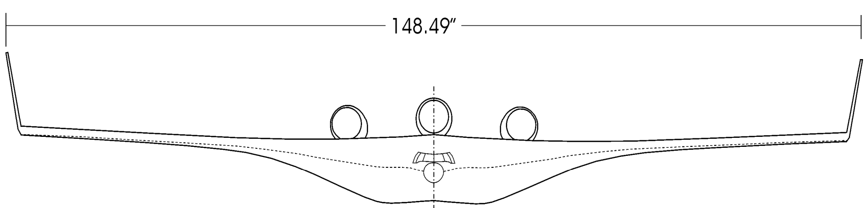
Old Dominion University

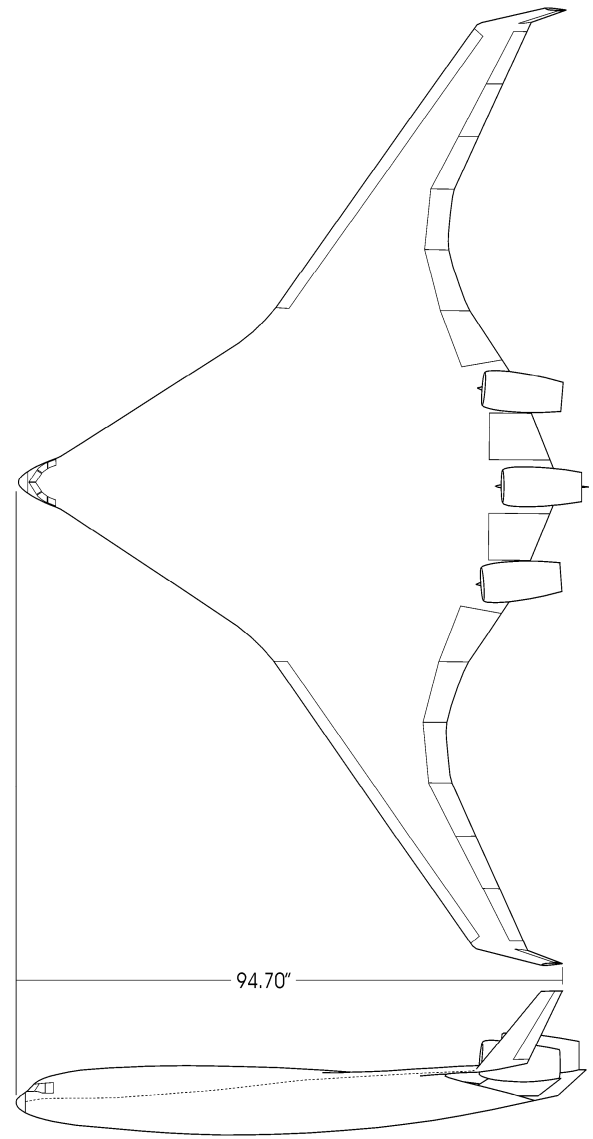
Low Speed Wind Tunnel Testing

- DOE methods are a natural fit for aerodynamic characterization
  - Typically large number of factors involved
    - Control surface deflections (at least 3-5 typically)
    - Attitude – 2 factors, sideslip and angle of attack
    - Power effects – thrust simulation
    - Landing gear up/down
    - Configuration changes (eg. Slat A or B)
  - Use of actuated control surfaces helps accommodate randomization
    - Loads at relatively low dynamic pressure will allow

BWB Case Study

- Potential Factors
  - 18 trailing edge mounted elevons
  - 3 pylon mounted
    engines with thrust simulation
  - 2 Leading edge slats
  - 2 Winglet rudders





OLD DOMINION UNIVERSITY

5% Scale Static Testing at the LFST

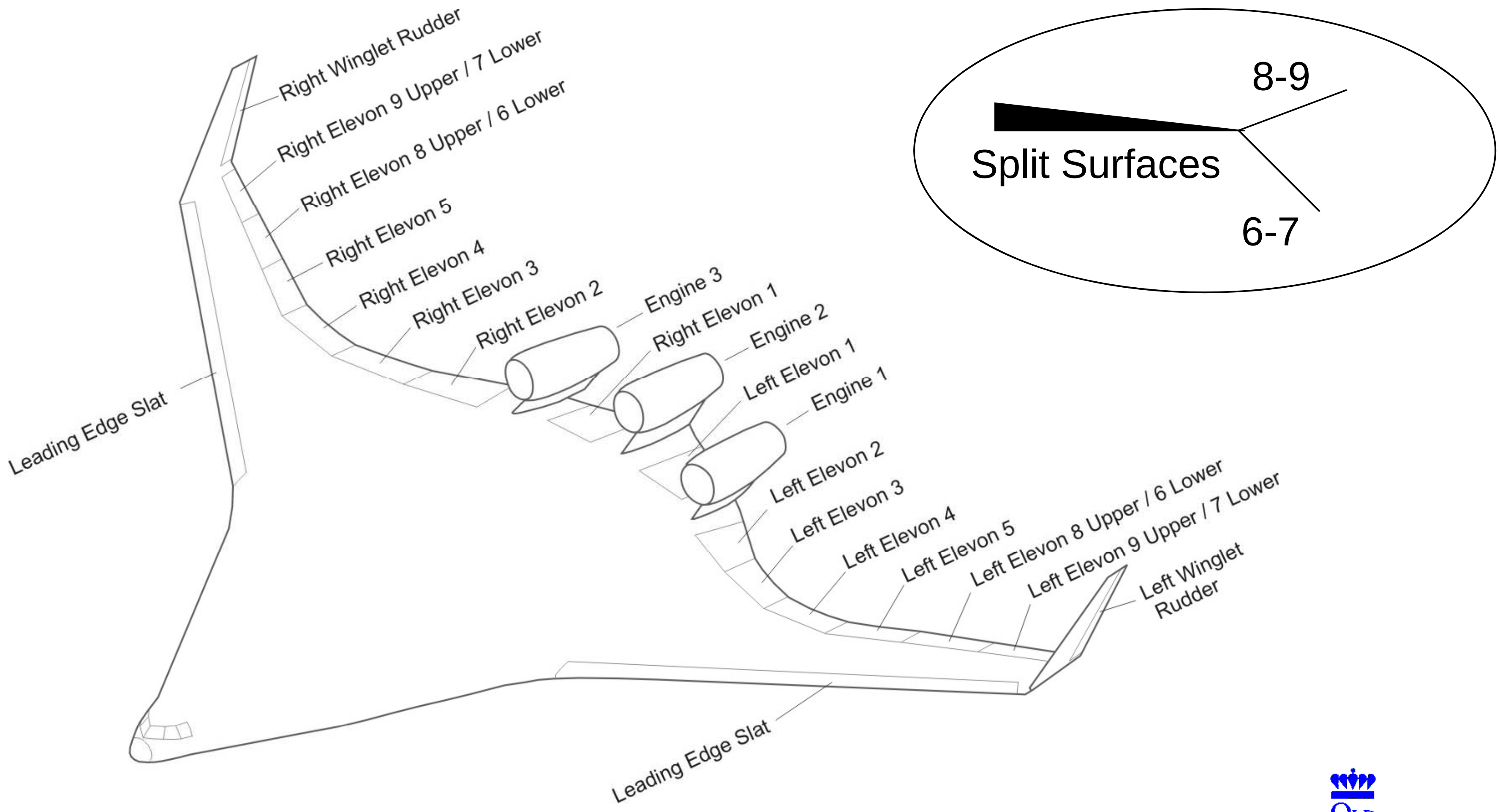
- Objectives
  - Aerodynamic characterization including
    - Static stability
    - Control power
- Sting supported
- Internal Balance
- Remotely actuated surfaces
- Air ejectors for thrust

OLD DOMINION UNIVERSITY

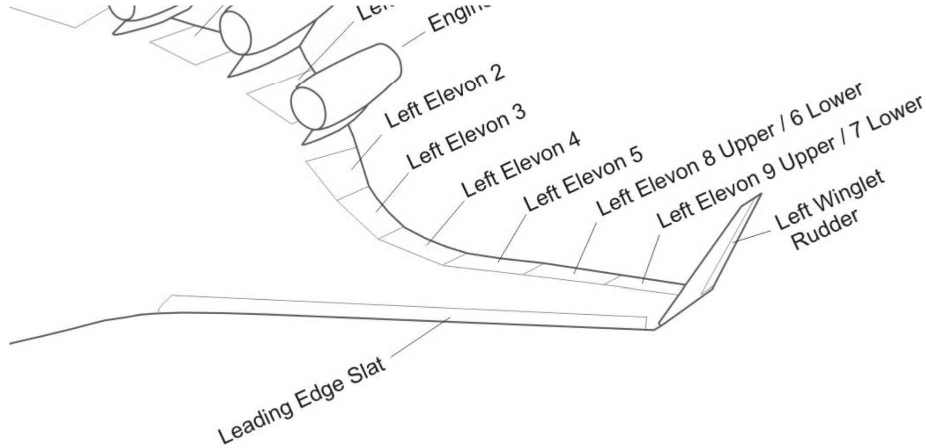
Stability and Control Test

- A subset of controls were chosen
  - Due to actual model limitations of ganged surfaces
  - Due to limited resources for the DOE test
- Look at response due to left wing control surfaces
  - Use full range of surface deflections
  - Look at one right wing surface
- Interested in possible interactions with adjacent surfaces
- Angle of attack chosen to bracket cruise conditions (angular orientation in vertical plane)
- Small sideslip range (angular orientation in horizontal plane)

BWB Model Control Surfaces



8 Factors Chosen



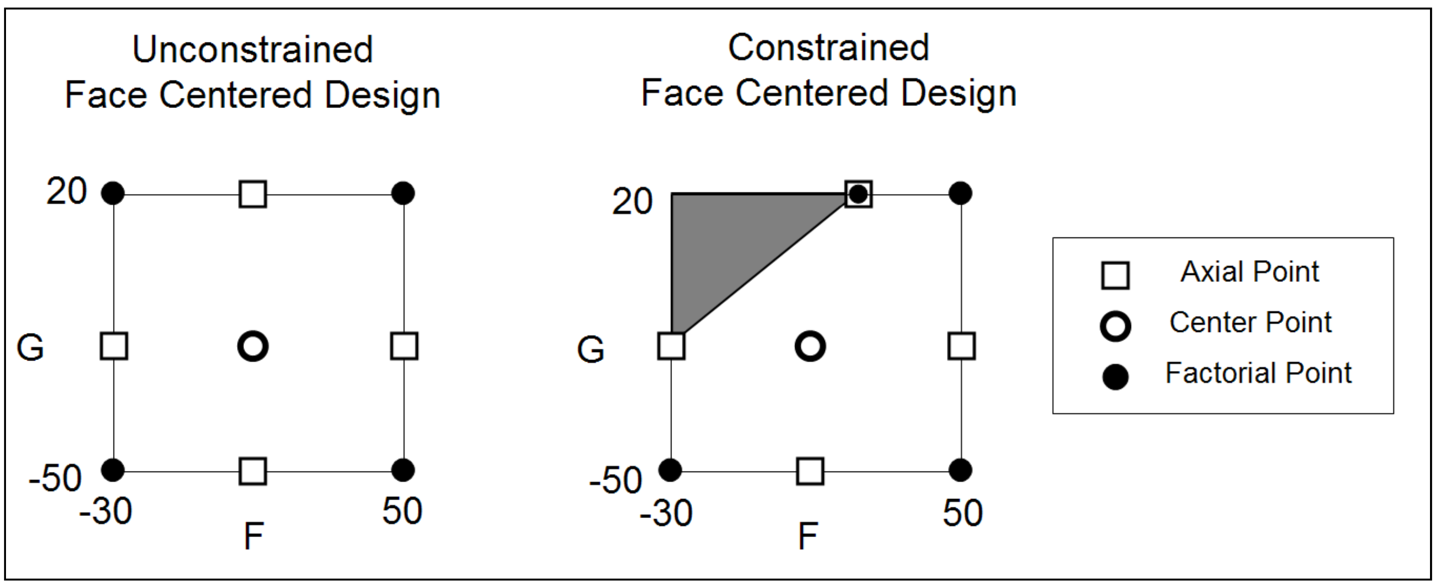
All are given in degrees

Factor	Factor ID	Low	Center	High	Constraints
α	A	4	7	10	none
β	B	-5	0	5	none
δ_{R25}	C	-30	-5	20	none
δ_{1}	D	-30	-5	20	none
δ_{L25}	E	-30	-5	20	none
δ_{L67}	F	-30	10	50	F - G > 5
δ_{L89}	G	-50	-15	20	F - G > 5
δ_{Lrud}	H	-20	5	30	none

Old Dominion University

BWB DOE Design Criteria

- Allow at least 3 levels for control surface set points
- Robust to control surface set point error
- Model order for force and moment responses
  - Pure quadratics, possible 2 F.I.'s and 3 F.I.'s
  - Can sequentially augment design by adding design points to build higher order model if required
- Design must accommodate constraints, modified CCD used



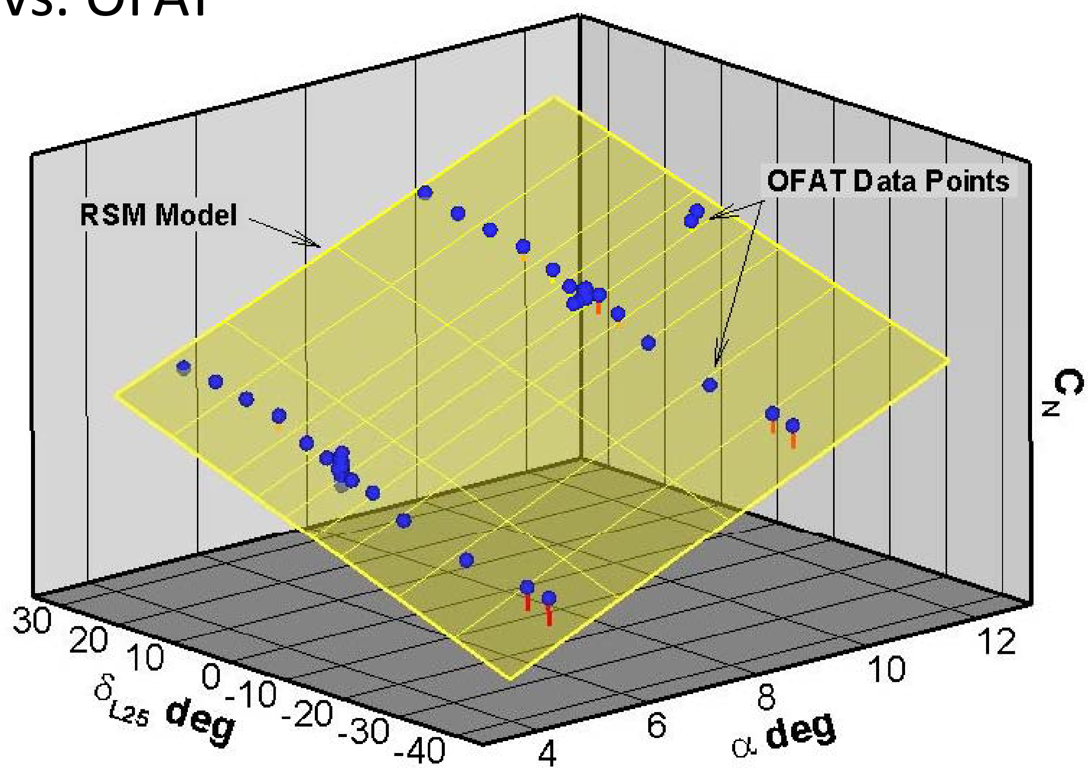
Experiment Details

- Fully automated test
  - Control surfaces actuated
  - Attitude control programmed
- Fully randomized test matrix
- Use 2 Blocks
  - Day 1: ½ Fraction factorial and ½ of center points
  - Day 2: Axials and remaining center points
  - Blocks to protect against "nuisance factors"

Old Dominion University

Results: Comparison to OFAT

- Example: Left Ganged Elevons 2-5 versus angle of attack, response is Normal force
- Half the data volume vs. OFAT
- Statistically justified uncertainty estimates
- Regression model as function of all factors
- All potential Interactions modeled

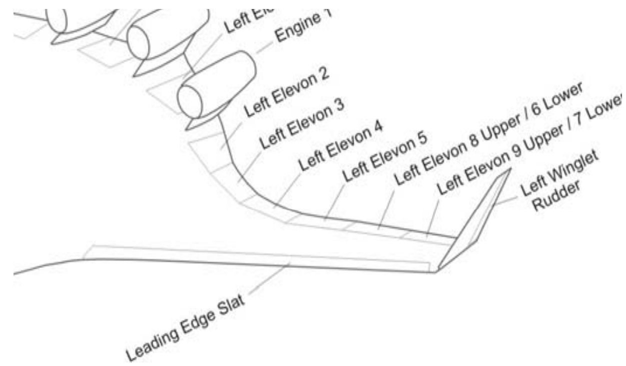


Optimization

- Complex control surface configurations present an interesting opportunity in that choices arise in allocation of surfaces to achieve specific control objectives
- DOE regression models include effects of all control surfaces and interactions – a natural for optimization
- Example (using Desirability approach)
  - Using any combination of control surfaces tested
  - Find <u>maximum yaw moment magnitude</u> with <u>minimum roll magnitude</u>

Example: Max Yaw for Min Roll

- Results show elevons 6-7 down and 8-9 up at the optimum, winglet rudder full deflection with 2-5 used to “trim”
- Roll moment (C_l) was nearly zero with yaw moment (C_n) 90 % of maximum



Solution	δ_{L25}	δ_{L67}	δ_{L89}	δ_{Lrud}	$C_{n\ (\%\ Best)}$	$C_l/C_{n\ (\%)}$	Desirability
1	-6.44	40.00	-33.98	30.00	90.11	0.842	0.823
2	-11.54	40.00	-35.85	30.00	100.00	80.686	0.806
3	-10.10	40.00	-13.39	29.93	73.61	0.000	0.789
4	2.69	-26.45	-35.83	30.00	71.63	0.001	0.785
5	2.50	-25.08	-35.85	29.97	70.61	0.188	0.783
6	-3.22	17.80	-35.85	30.00	69.77	-0.003	0.781
7	1.45	-17.31	-35.85	29.88	65.91	-0.003	0.773

OLD DOMINION UNIVERSITY

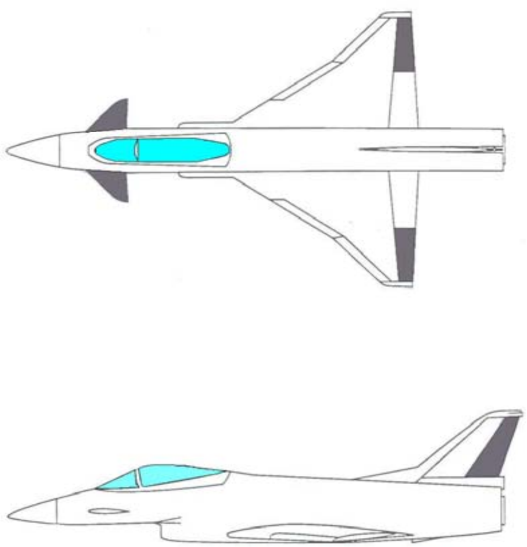
Summary

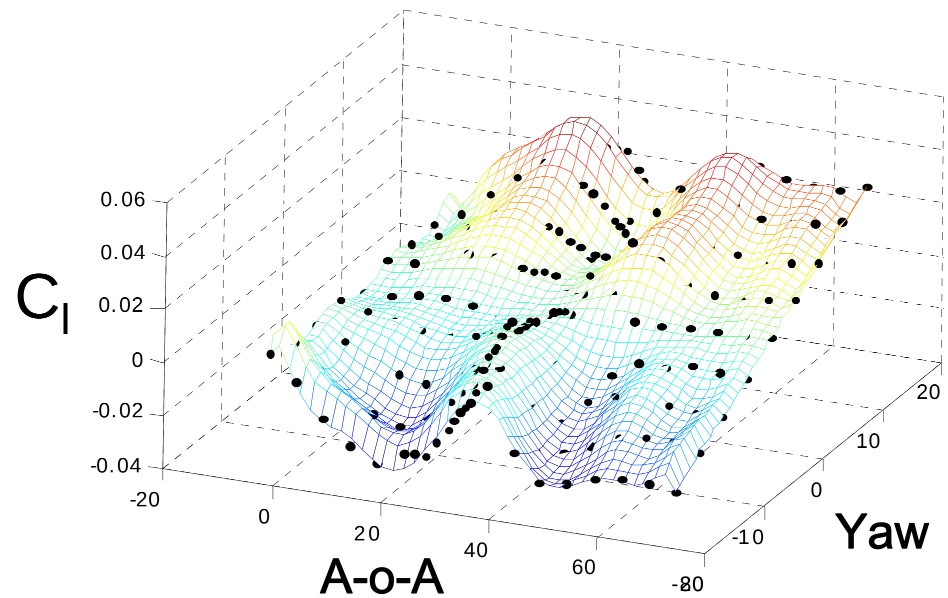
- Regression models developed for all aerodynamic forces and moments as a function of 8 factors
- Uncertainty estimates including effects of exercising all factors
  - Researchers initially felt these estimates were high after comparing to OFAT estimates
  - Residual analysis revealed problems with actuated surfaces reaching set points that were undiagnosed in preceding 2-week long OFAT investigation
- Identification of interactions on new configuration
- Optimization opportunities

OLD DOMINION UNIVERSITY

Low Speed Wind Tunnel Test Challenges: Highly Non-Linear Regions

- How to model with low order polynomials
  - Rolling moment of X-31 high-performance aircraft example

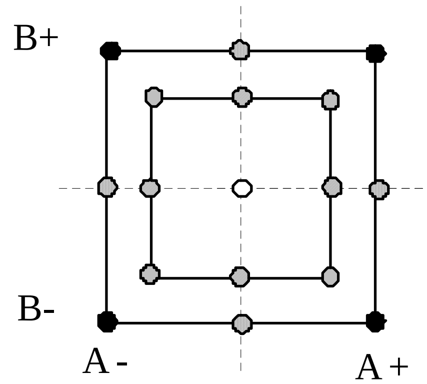




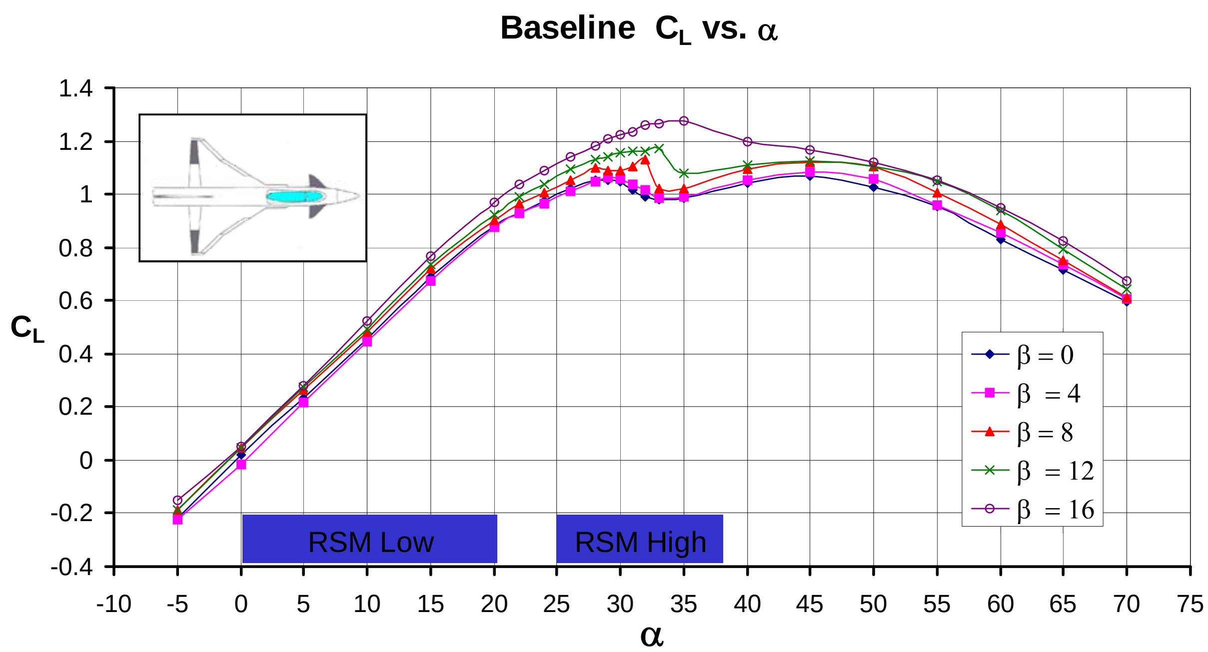
OLD DOMINION UNIVERSITY

Highly Non-Linear Regions

- Exploratory study required
- Higher order hybrid designs valuable
- Design space may be broken into subspaces
  - Issue often cited is how to handle adjacent spaces at the borders
    - Fairing functions
    - Final use is typically lookup tables



Nested FCD Hybrid



LSWT Challenges: Hard-To-Change Factors

- Factor choices that require replacement of parts may be impractical for a fully randomized design
  - One solution is to run a split-plot design, introduces multiple error sources
  - Analysis is complicated but now supported by commercial software
- Example MAV, wingtip height is hard-to-change (a whole plot factor), A-o-A and Yaw are easy to change (subplot factors)

Standard Order	*Run Order*	*Wingtip Height*	*AoA*	*Yaw Angle*
20	1	3	6	0
17	2	3	14	0
19	3	3	6	10
16	4	3	-2	0
18	5	3	6	-10
14	6	4	14	10
12	7	4	-2	10
11	8	4	-2	-10
15	9	4	6	0
13	10	4	14	-10
2	11	2	-2	10
1	12	2	-2	-10
3	13	2	14	-10
5	14	2	6	0
4	15	2	14	10
10	16	3	6	0
9	17	3	6	10
6	18	3	-2	0
8	19	3	6	-10
7	20	3	14	0

OLD DOMINION UNIVERSITY

High Speed Wind Tunnel Testing Challenges

- Traditional DOE methods rely on randomization, difficult for:
  - Factors of Reynolds and Mach number
    - Resource intensive to randomly set combinations
  - Control surface deflections as factors
    - Loads are often too high to use actuators
    - fixed brackets are common (and make for hard to change factors)
  - Factors sideslip and angle of attack
    - Often possible to automate
- Again, one solution is a split plot design structure

Old Dominion University

Challenge: Training

- How do we train the engineering staff in use of DOE ?
  - Experience with training at AEDC
    - Culture change - observations
      - It’s a steep learning curve for those that have never studied statistics or regression
      - Focus has historically been on rapid acquisition of high precision data - **not** empirical model adequacy/quality
      - Sequential experimentation is often viewed as unnecessary
      - Fractional factorials felt to leave too much up to chance
    - Aversion to randomization
      - Many hard to change factors in AEDC facilities works against easy adoption of DOE
      - Reliable, low noise experimental facilities often breed cavalier attitudes: “Protect against what lurking variable – we don’t have those”

Old Dominion University

Some References

"Adapting Second-Order Response Surface Designs to Specific Needs," Simpson, J.; Landman, D.; Giroux, R.; Zeisset, M.; Hall, B.; Rhew, R.; *Quality and Reliability Engineering International,* Vol. 24, No. 3, pp. 331-349, April 2008.

"Response Surface Methods for Efficient Complex Aircraft Configuration Aerodynamic Characterization," Landman, D.; Simpson, J.; Vicroy, D.; and Parker, P.; AIAA Journal of Aircraft, Vol. 44, No. 4, pp. 1189-1195, July-August 2007.

"Hybrid Design for Aircraft Wind-Tunnel Testing using Response Surface Methodologies," Landman, D. et al., AIAA Journal of Aircraft, Vol. 44, No. 4, pp. 1214-1221, July-August 2007.

Low-Speed Wind Tunnel Testing via Designed Experiments: Challenges and Ways Forward," Simpson, J.; Landman, D.; AIAA paper 2008-1664, presented at the 2008 AIAA USAF T and E Days, Los Angeles, CA February 5-7, 2008.

"Development of Tandem Wing MAV Wind Tunnel Test Plan Using Experimental Design Techniques," English, T.; Landman, D.; Simpson, J.; presented at the INFORMS International Conference, Rio Grande, Puerto Rico, July 8-11, 2007.

"A Wind Tunnel External Balance Calibration using Design of Experiments," Landman, D.; Simpson, J.; AIAA paper 2007-1604, presented at the AIAA US Air Force T and E Days, February 13-15, 2007.

Old Dominion University

Contact

Drew Landman, P.E, Ph.D.
Professor
Department of Mechanical and Aerospace Engineering
Old Dominion University
Norfolk, VA 23529
dlandman@odu.edu
office 757-683-6008
http://eng.odu.edu/mae/directory/dlandman.shtml

Opportunities for Statistical Collaboration with NASA: Some Personal Reflections

Geoff Vining

Virginia Tech

Outline

- Some Personal Background
- My Background with NASA
  - LaRC
  - NESC
- Reflections
- Opportunities

Some Personal Background

- University of Florida
  - Typical Academic Career
  - Large Amount of Consulting
    - Through the University
    - Outside the University
  - Gates Aerospace Batteries: LEO – ISS Batteries

Some Personal Background

- Virginia Tech
  - Department Head
    - “Seven years of college down the drain.”
    - Corporate Partners Program
  - Consulting as a Full Professor
    - Pratt & Whitney
    - NASA LaRC
    - NESC
    - DoD
  - Mostly, Design of Experiments

My Background with NASA

- LaRC
  - Dr. Parker Was My Ph.D. Student
  - First Collaborations Were with the Atmospheric Sciences
  - Recently, Working on Calibration
    - ARES
    - AirSTAR
    - Force Measurement Systems
    - A.G. Davis System
  - Two Publications to Date

My Background with NASA

- NESC – COPV
  - Ultimate Question: Probability of Failure for a COPV at Use Conditions at Specific Time in Future
  - Design of Experiments for Reliability Data
    - DOE People Rarely Understand Lifetime Data
    - Reliability Experts Rarely Understand the Nuances of DOE
  - Initial Work: Testing “Strands”
  - Some Preliminary Work: Vessels

Reflections

- My Primary Involvement: Technical Expert
  - Some Work Helping to Define the Problem
    - Usually, Initial Work Already Done on Problem Definition
    - My Job: Refine and Clarify as Needed
  - Project Manager
    - Supervise Graduate Students
    - Administration
    - Provide Technical Guidance – Both for VT and NASA
    - "The Right Tool for the Right Job."
  - Major Contribution: Big Picture

Reflections

- As a Rule, NASA Under-Utilizes Statistics and Statisticians
  - NASA Employs Very Few True Statisticians
  - Most Engineers/Scientists Have Limited Command of Proper Statistical Procedures
- A Large Amount of Mathematical Modeling, but Very Little Statistical Modeling
- Combination Can Lead to Questionable Practices

Reflections

- Very Little Statistical Thinking within NASA
  - All Work Occurs in Systems of Interconnected Processes
  - Variation Exists in All Processes
  - Keys to Success: Understanding and Reducing Variation
- Many NASA Statisticians Are Pure Data Analysts, Not Scientific Collaborators/Leaders
- Strong Agency Need for True Statistical Engineering and Statistical Leadership

Reflections

- Statistical Engineering and My Involvement
  - My Work with Dr. Parker: Statistical Engineering
    - Problem Selection
    - Decisions about Tactical Deployment
    - Focus on Big Picture and Value Added
  - Otherwise: Traditional Consulting
    - One Step Below Tactical
    - Focus on Reasonable Solutions to Specific Problems
    - Teaching Graduate Students to Consult
    - "Tool" in the Nascent Statistical Engineering Effort

Reflections

- Benefits to Date: Virginia Tech
  - Teaching Graduate Students to Consult with Practicing Engineers and Scientists on Real Problems
  - Students See: Good, Bad, Ugly
  - Publishing Research Papers Based on Real Engineering and Science Problems
  - Pride in Helping a Distinguished Gov. Agency

Reflections

- Benefits: NASA
  - Technical Support Not Available within the Agency
    - Limited Number of Statisticians within the Agency
    - Academics Often are the Leading Experts in the Field
  - Pipeline for Hiring Ph.D. Statisticians
    - Already Familiar with the NASA, Its Mission and Culture
    - Real Practical Experience with Agency Problems
  - Standard Benefits from University/Agency Collaboration

Opportunities

- Can Achieve Benefits Far Above Standard University/Agency Collaboration!
- Implement on a Broader Scale True Statistical Engineering
- Biggest Initial Contributions:
  – Statistical Thinking/Clear Problem Definition
  – Sound Structured Approaches to Solving Problems
- At Least Initially, Take Advantage of Academic and Professional Expertise

Opportunities

- NASA Technical Leadership Needs to Provide the Strategic Vision
  - Set Agency – Wide Goals and Objectives
  - Recruit Appropriate Personnel from within NASA
  - Manage the Entire Process
- "NASA Statistical Engineering Group"
  - Develop Tactical Plans to Achieve Strategic Goals
  - Manage Specific Projects Selected by Leadership

Opportunities

- NASA Deals Daily with Highly Complex Problems
- A Large Portion of These Problems Have Significant Statistical Components
  - Some Cases, Understood
  - Far Too Many Cases, Not Understood!
- Statistical Engineering Provides Tactical Deployment of Sound Statistical Practices to Support the Engineering/Scientific Method

Opportunities

- Complex Problems Require Appropriate Solutions
  - Team Approach to Solutions
  - Clear and Precise Problem Definitions
  - Avoid Errors of the Third Type!
  - Systems Thinking
  - Understanding Sources of Variation
  - Appropriate Data
  - Tactical Deployment of Analytics
- Statistical Engineering Is an Appropriate Approach!

Opportunities

- Implementing Statistical Engineering Is a Journey
- Academia and the Profession Can Provide Excellent Guides
- The Process Leads to Better Science and Engineering, Which Is Core to NASA's Mission

Design of Experiments in Measurement System Characterization and Uncertainty

Tom Johnson

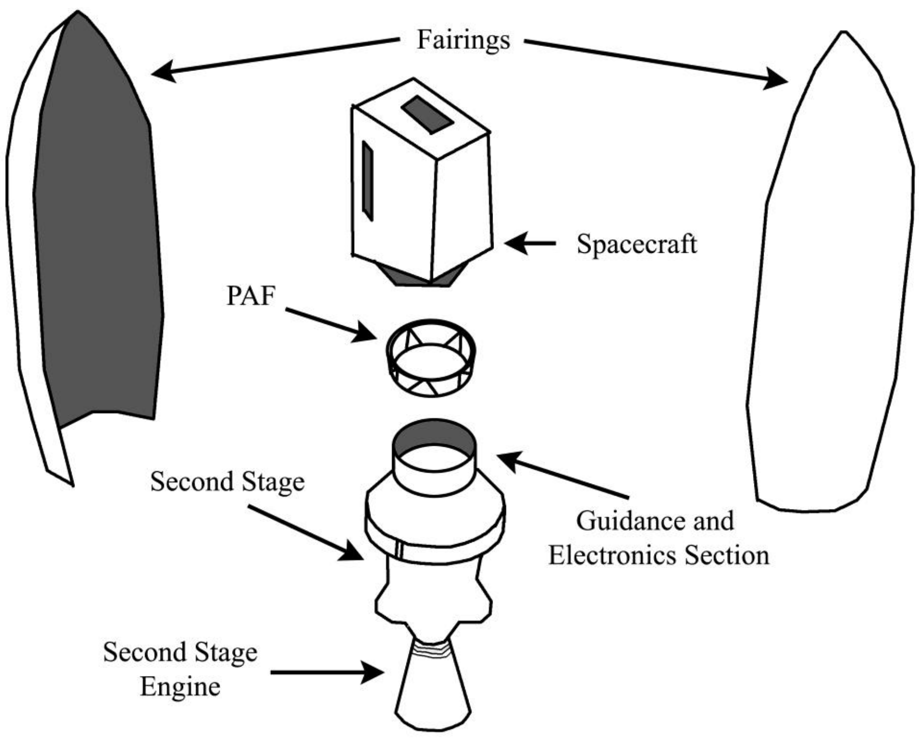
03/22/11

Outline

1. In-Flight Force Measurement Method
2. Non-monolithic Calibration Design
3. Variable Acceleration Calibration System
4. Center of Gravity Determination Method

In-Flight Force Measurement Method

- What is the problem?
  - There is a requirement for an alternative method to monitor in-flight loads experienced by a particular spacecraft during launch
- Why does it matter?
  - To properly understand the physics of the problem
  - To ensure the safety of the spacecraft
  - to achieve a successful launch
- Who does it matter to?
  - Engineers, researchers
  - Customers using the delta II rocket
  - Boeing to maintain a reliable track record
- Project Objective
  - Monitor loads exerted on spacecraft during launch (pre-determined)
  - Adapt a structural piece of a Boeing delta II rocket, called a Payload Attachment Fitting (PAF), into multi-component force transducer (also pre-determined)
- What is the PAF?



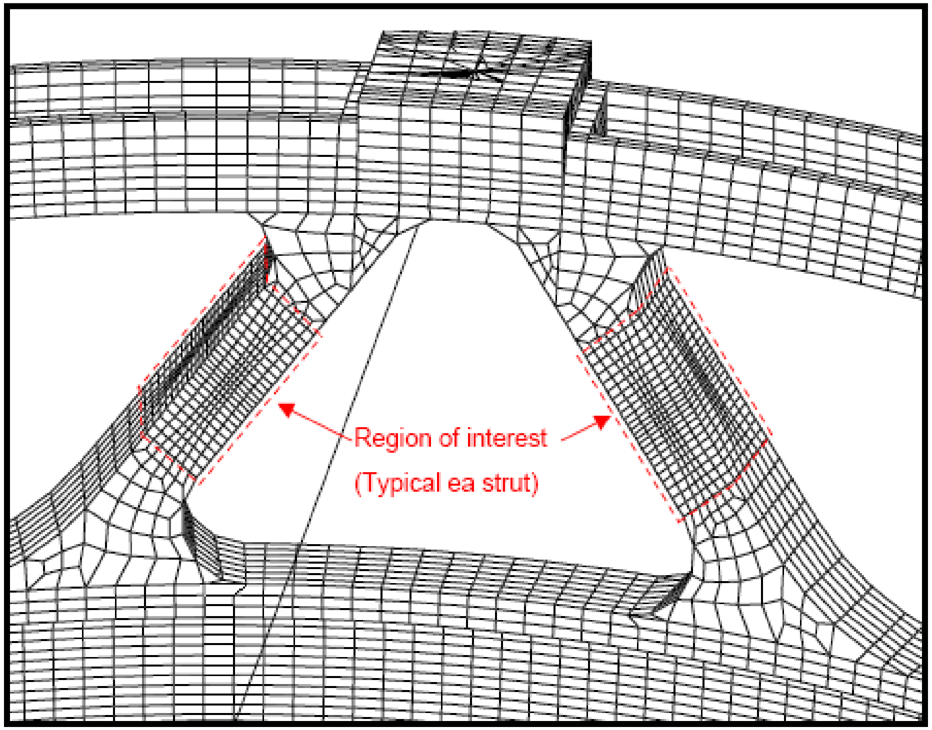
"A Multi-Component Force Transducer Design from an Existing Rocket Payload Attachment Fitting," Johnson,T.; Landman,D; Parker, P.; AIAA-2009-1716, AIAA USAF T and E Days 2009, Albuquerque, NM, February 10-12, 2009.

In-Flight Force Measurement Method

- *Proposed Solution Framework*
  1. First, optimize strain gauge locations on the PAF using computer simulation (Finite Element Analysis)
  2. Instrument PAF according to strain gauge location optimization study
  3. Perform a ground based calibration
  4. Use in-flight data with the calibration models to obtain in-flight forces

- *Strain gauge optimization method*
  - Objective: determine strain gauge locations that maximize the sensitivity of the reading for a given force component, while minimizing interactions effects due to other forces.
  - Using design of experiment, a factorial design was run to model strain as a function of applied forces at each element in the FEA model.
  - Factors (6): applied loads
  - Responses (>10,000): strain at each element

$$y_n = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j}\sum \beta_{ij} x_i x_j + \epsilon$$



In-Flight Force Measurement Method

- optimization method (continued)
  - The next step of the problem is to find combinations of 4 strain locations that maximizes sensitivity while minimizing interaction effects
  - Since Wheatstone bridges require 4 gauges
  - Numerical search method used to find best combination of gauges for each model
  - 4 gauges for each component resulted in 24 gauge locations total
- Optimization results
  - Approximate 10 lb resolution
- Proposed next steps
  - Instrument the PAF
  - Perform ground based calibration
- Conclusion
  - A completely unique method for determining gauge location methods was demonstrated
  - Design of experiments was used to make efficient use of computational resources

Main effects coefficients

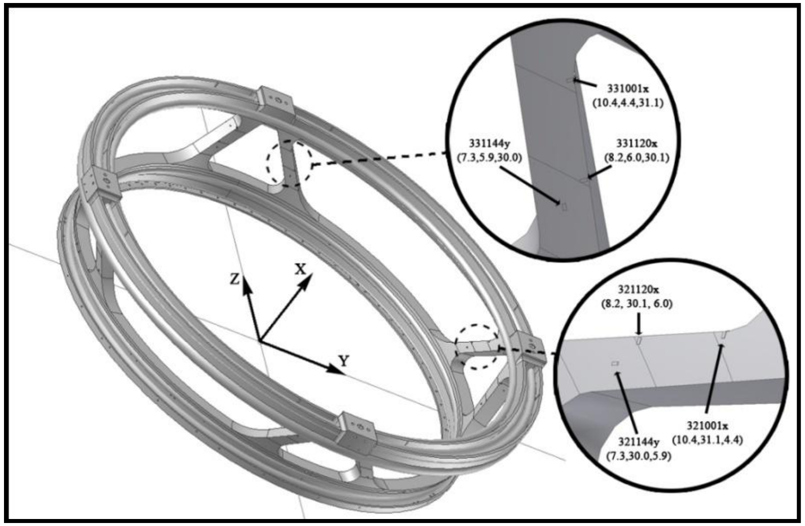
Applied loads

2FI's

$$y_n = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum\sum_{i<j} \beta_{ij} x_i x_j + \epsilon$$

Strain at a finite element

n – varies from 1 to 12,302 , 1 for each element of the finite element model



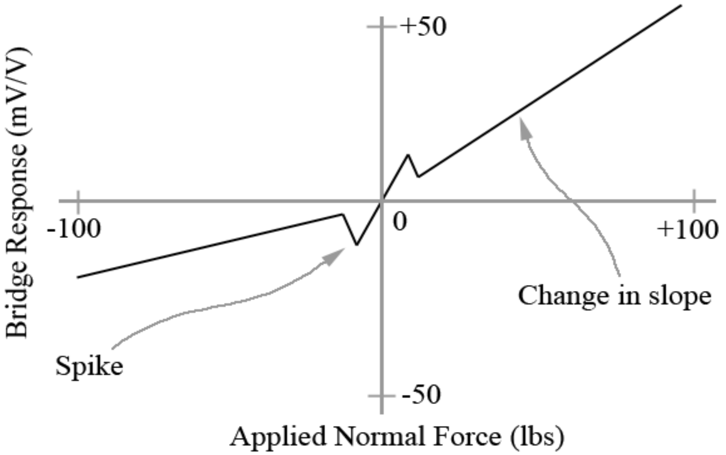
Outline

1. In-Flight Force Measurement Method
2. Non-monolithic Calibration Design
3. Variable Acceleration Calibration System
4. Center of Gravity Determination Method

Non-Monolithic Calibration Design

- What is a force balance?
- What is the problem?
  - It is nationally recognized that current methods used to model non-monolithic force balances is inadequate
- Why does it matter?
  - There are many non-monolithic balances currently being used to characterize the performance of tomorrows space vehicles
  - The performance of future missions relies on the accuracy of wind tunnel tests
- Who does it matter to?
  - Force measurement community
  - AIAA reccommended standard calibration practices document
  - Project leaders
- Project Objective
  - Demonstrate shortcomings of current recommended procedure
  - Propose alternative solutions

- What is a non-monolithic balance?



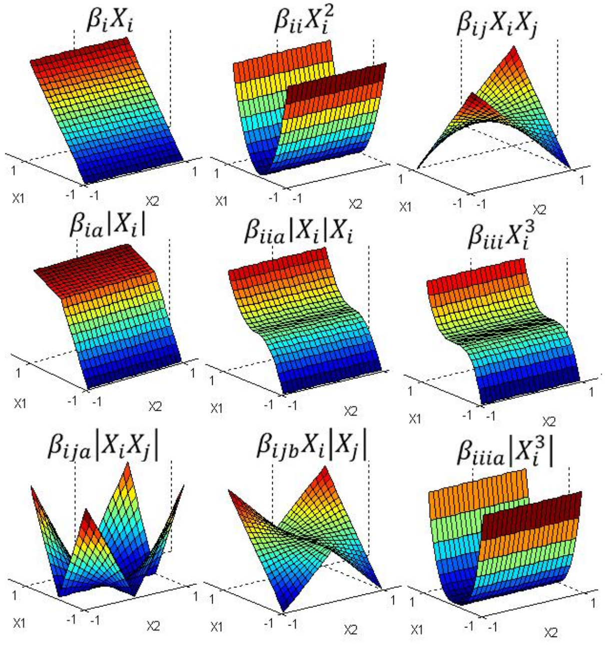
Johnson, T. H., Parker, P.A., Landman, D., "Calibration Modeling of Nonmonolithic Wind-Tunnel Force Balances," AIAA-46356-110 , AIAA Journal of Aircraft, Vol. 47, No. 6, Nov-Dec 2010.

Non-Monolithic Calibration Design

- Current standard procedure recommends using the model shown below
- Takes a heavily parameterized approach
- Includes absolute value terms to model asymmetry in the response
- The problem with the model is that it is over parameterized

$$R_i = a_i + \sum_{j=1}^{n} b1_{i,j}F_j + \sum_{j=1}^{n} b2_{i,j}|F_j| + \sum_{j=1}^{n} c1_{i,j}F_j^2$$
$$+ \sum_{j=1}^{n} c2_{i,j}F_j|F_j| + \sum_{j=1}^{n}\sum_{k=j+1}^{n} c3_{i,j,k}F_jF_k$$
$$+ \sum_{j=1}^{n}\sum_{k=j+1}^{n} c4_{i,j,k}|F_jF_k| + \sum_{j=1}^{n}\sum_{k=j+1}^{n} c5_{i,j,k}\,F_j|F_k|$$
$$+ \sum_{j=1}^{n}\sum_{k=j+1}^{n} c6_{i,j,k}\,|F_j|F_k + \sum_{j=1}^{n} d1_{i,j}\,F_j^3$$
$$+ \sum_{j=1}^{n} d1_{i,j}\,|F_j^3|$$

- Certain parameters in the model should not co-exist no matter what experimental design is used
- The figure to the bottom right shows response surfaces of various effects from the model
- Variance Inflation Factors are used to show multi-collinearity between model parameters



Johnson, T. H., Parker, P.A., Landman, D., "Calibration Modeling of Nonmonolithic Wind-Tunnel Force Balances," AIAA-46356-110 , AIAA Journal of Aircraft, Vol. 47, No. 6, Nov-Dec 2010.

Non-Monolithic Calibration Design

Eq #	Model	Parameters	Design	# of Runs
1	Independent	28	2 CCDs	128
2	Cubic	55	Draper	228
3	Absolute Value	34	Draper	228
4	Indicator Variable	28	2 CCDs	128

$$R_i = a_i + \sum_{j=1}^{n} b1_{i,j}F_j + \sum_{j=1}^{n} c1_{i,j}F_j^2 + \sum_{j=1}^{n}\sum_{k=j+1}^{n} c3_{i,j,k}F_jF_k \quad (1)$$

$$R_i = a_i + \sum_{j=1}^{n} b1_{i,j}F_j + \sum_{j=1}^{n} b2_{i,j}F_j^2 + \sum_{j=1}^{n}\sum_{k=j+1}^{n} b3_{i,j,k}F_jF_k + \sum_{m}^{n}\sum_{j=m+1}^{n}\sum_{k=m+j+1}^{n} b4_{i,m,j,k}F_mF_jF_k + \sum_{j=1}^{n} b5_{i,j}F_j^3 \quad (2)$$

$$R_i = a_i + \sum_{j=1}^{n} b1_{i,j}F_j + \sum_{j=1}^{n} b2_{i,j}|F_j| + \sum_{j=1}^{n} b3_{i,j}F_j|F_j| + \sum_{j=1}^{n}\sum_{k=j+1}^{n} b4_{i,j,k}F_jF_k \quad (3)$$

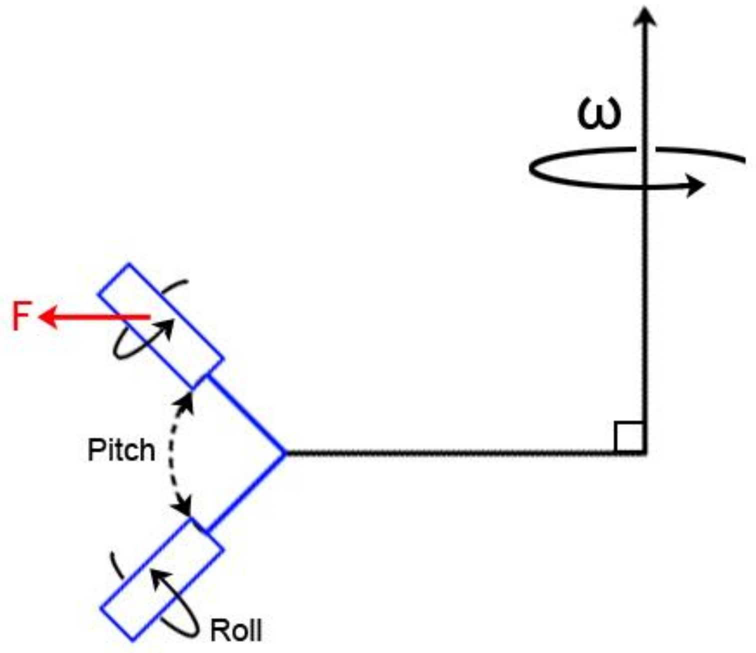
$$R_i = a_i + \sum_{j=1}^{n} b1_{i,j}F_j + \sum_{j=1}^{n} \Psi_{i,j}Z_{i,j}F_j + \sum_{j=1}^{n}\sum_{k=j+1}^{n} b4_{i,j,k}F_jF_k \quad (4)$$

Outline

1. In-Flight Force Measurement Method
2. Non-monolithic Calibration Design
3. Variable Acceleration Calibration System
4. Center of Gravity Determination Method

Variable Acceleration Calibration System

- What is a calibration system?
- What is the problem?
  - Calibration of large-scale internal wind tunnel force balances is expensive and inefficient.
- Why does it matter?
  - Large balances are needed for experimentation in NASA wind tunnels
  - Needed for full-scale wind tunnel test or semi-span tests
- Who does it matter to?
  - Force balance community, wind tunnel researchers, project engineers
- Project Objective
  - Design, fabricate and test two proof of concept variable acceleration calibration systems
  - Verify the applied load accuracy is within the predicted bounds
  - Propose next stage of system development
- Novelty of approach
  - Applies centripetal and gravitational force to a balance to reduce the time associated with moving weight



Variable Acceleration Calibration System

1. Design Experiment

2. Physics Model

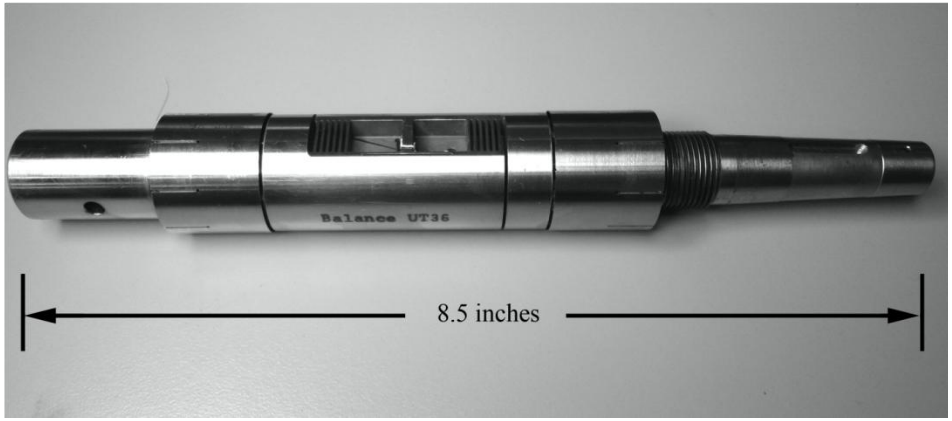
3. Mechanical Design

4. Run Experiment

5. Verification

- 3 Factors: Normal Force (NF) (lbs), Axial Force (AF) (lbs), Pitching Moment (PM) (in-lbs)
- 3 Responses: NF (volts), AF (volts), PM (volts)
- Fully replicated central composite design in two blocks

	NF (lbs)	AF (lbs)	PM (in-lbs)
Balance Design Loads	100	60	800
Calibration Loads	30	20	120



Variable Acceleration Calibration System

1. Design Experiment

2. Physics Model

3. Mechanical Design

4. Run Experiment

5. Verification

- Develop a physics-based prediction model to determine independent variable settings required to apply loads

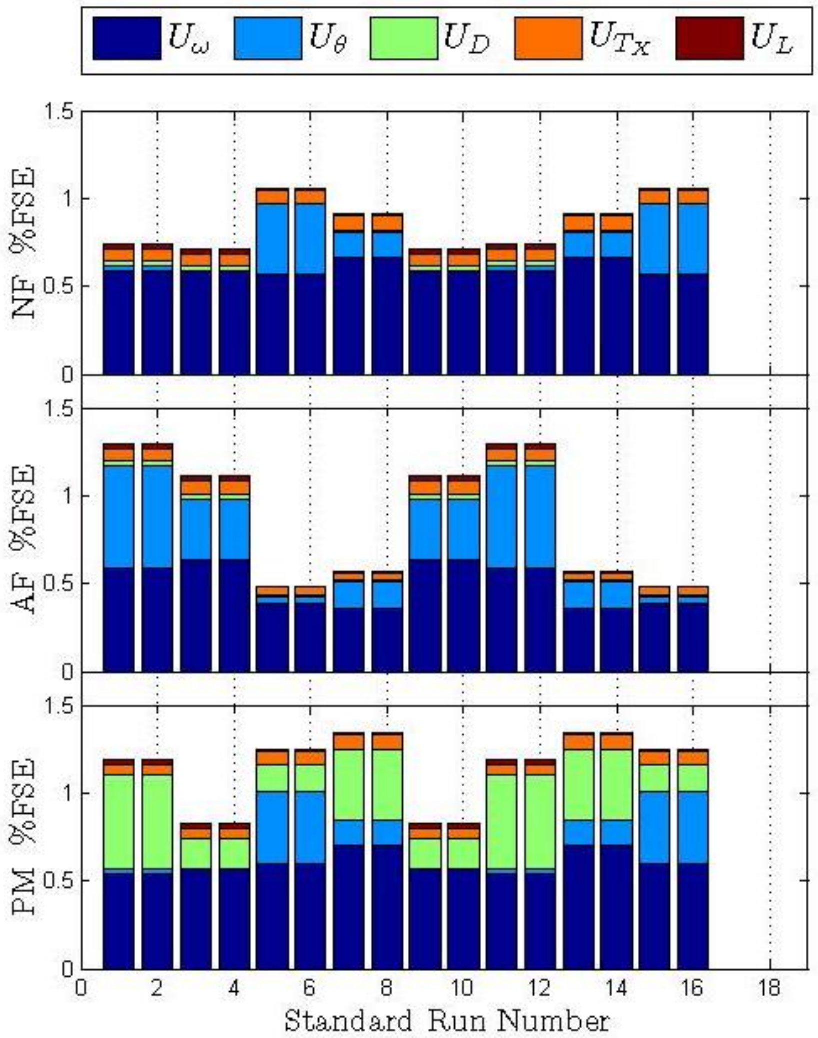
$$\boldsymbol{X} = [\omega \quad R \quad D \quad T_x \quad L \quad \theta \quad \phi \quad \alpha \quad m]$$

- Determine predicted uncertainty using propagation of uncertainty analysis

$$U_{pred,NF} = \sqrt{\sum_{j=1}^{9} \left(\frac{\partial NF}{\partial X_j} u_{X_j} \right)^2}$$

$$\boldsymbol{u_X} = [u_\omega \quad u_R \quad u_D \quad u_{T_x} \quad u_L \quad u_\theta \quad u_\phi \quad u_\alpha \quad u_m]$$

- Predicted independent variable uncertainty contributions shown to right for each run in calibration experiment



Variable Acceleration Calibration System

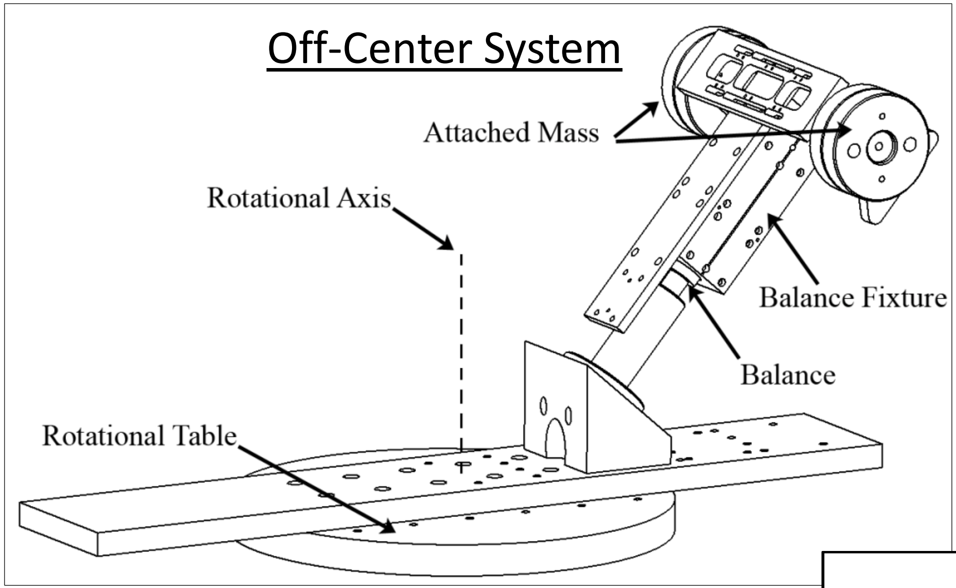
1. Design Experiment

2. Physics Model

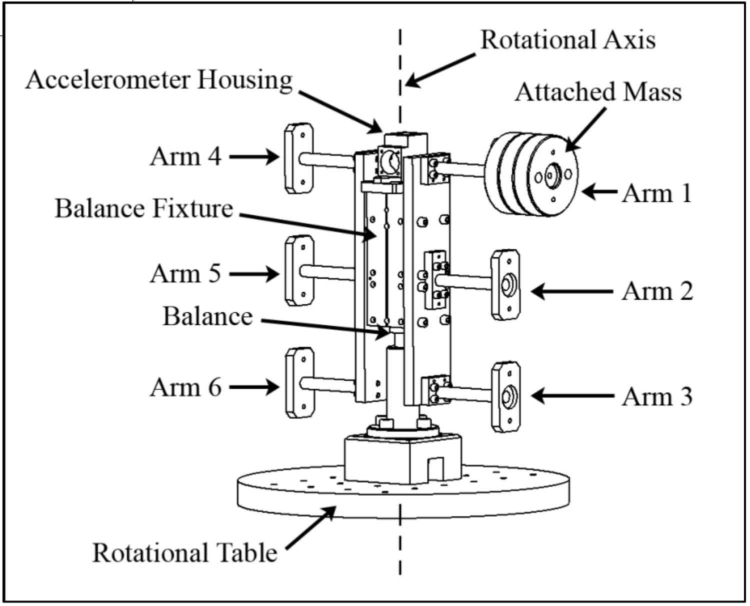
3. Mechanical Design

4. Run Experiment

5. Verification



Centered System



Variable Acceleration Calibration System

1. Design Experiment

2. Physics Model

3. Mechanical Design

4. Run Experiment

5. Verification

Variable Acceleration Calibration System

1. Design Experiment

2. Physics Model

3. Mechanical Design

4. Run Experiment

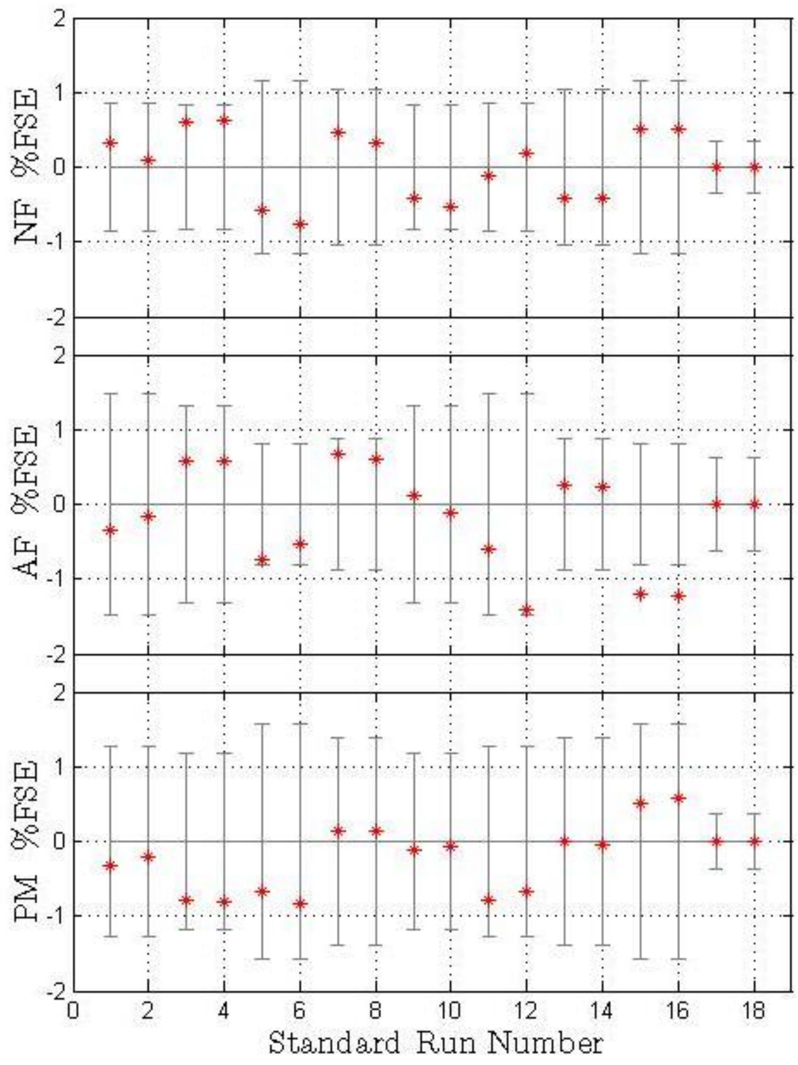
5. Verification

- Verify applied load error is within predicted uncertainty

- Predicted uncertainty contains
  - Uncertainty predicted using propagation analysis
  - Balance measurement uncertainty
  - Pure error (noise)

- Applied Load error is the physics model predicted loads minus balance measured loads (shown in red in figure to right)

- Residual Analysis
  - plot applied load error vs. independent variable
  - Plot pure error vs. independent variable graphs



Outline

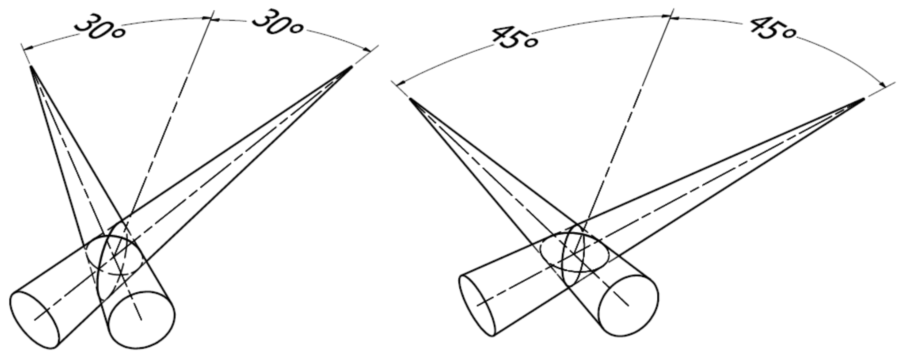
1. In-Flight Force Measurement Method
2. Non-monolithic Calibration Design
3. Variable Acceleration Calibration System
4. Center of Gravity Determination Method

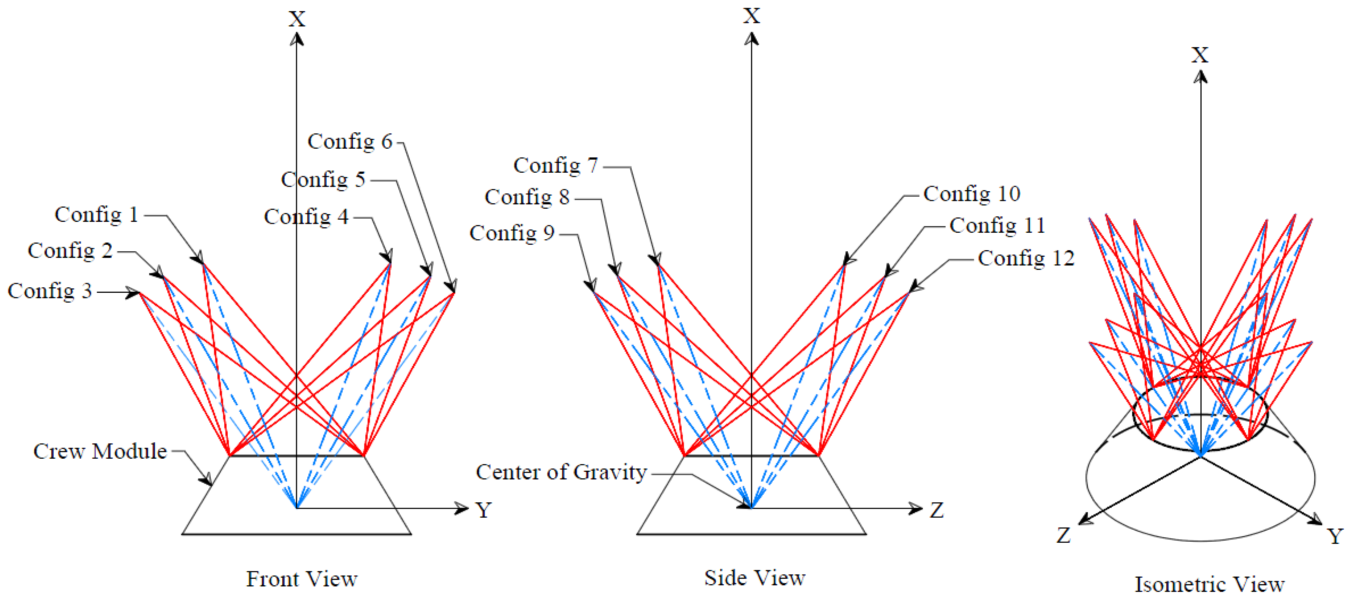
Center of Gravity Determination Method

- What is the problem?
  - CAD center of gravity results are not perfect. Experimental verification is often required for space vehicles.
- Why does it matter?
  - Center of gravity info is critical to guidance, navigation and control of spacecrafts.
- Objective
  - Create a inexpensive and efficient experimental method to determine the center of gravity of a space vehicle
  - Provide repeatable and statistically defendable results
- How does it work?
  - Geometry measurements are recorded for multiple test article hang configurations
  - A gravity vector is projected from the hang vertex in each configuration
  - The center of gravity is found by determining the “intersection point” of the multiple gravity vectors.

Center of Gravity Determination Method

- CG Method by Tom Jones, NASA LaRC
- My Contribution:
  - Help mature a concept and define the uncertainty
- Reduce uncertainty by reducing prediction uncertainy, quantifying experimental uncertainty
  - Proposed new hang angle to reduce intersection volume uncertainty
  - Orthogonal intersection reduces volume of uncertainty by 15% (compared to 60 deg intersection)



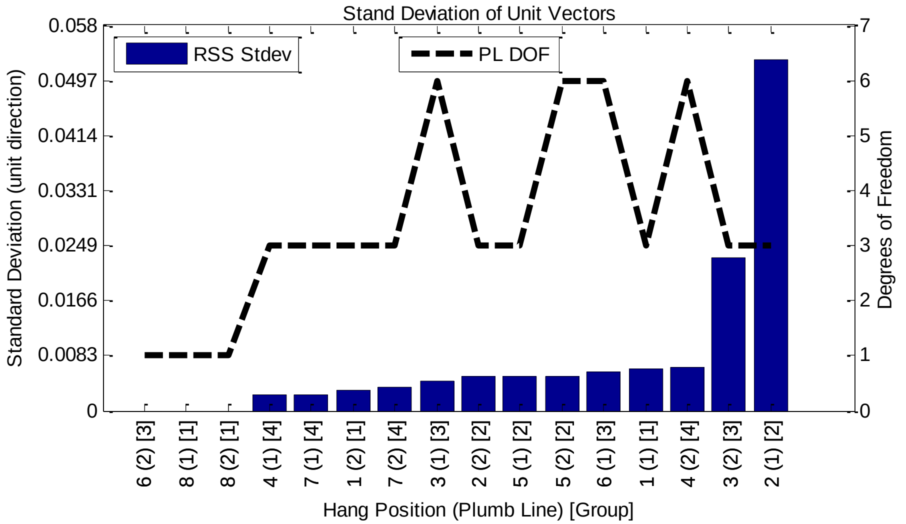


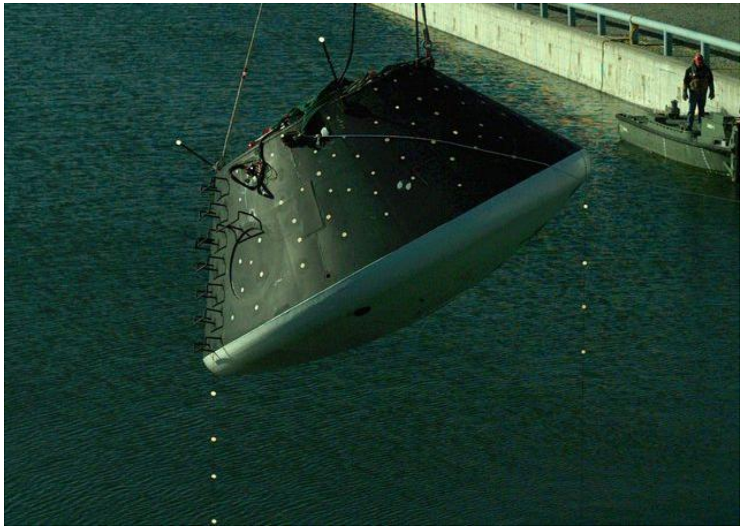
Center of Gravity Determination Method

- Gravity vector construction contains uncertainty due to wind and water effects
- Each gravity line was formed using 2-5 photogrammetry targets
- Pairs of targets within each gravity line were use to construct gravity direction

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

- The standard deviation was calculated for each line in each hang configuration to determine which lines had the most noise
- The lines with the least amount of noise were used for the CG calculation

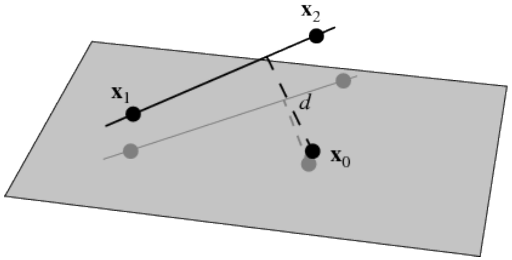




Center of Gravity Determination Method

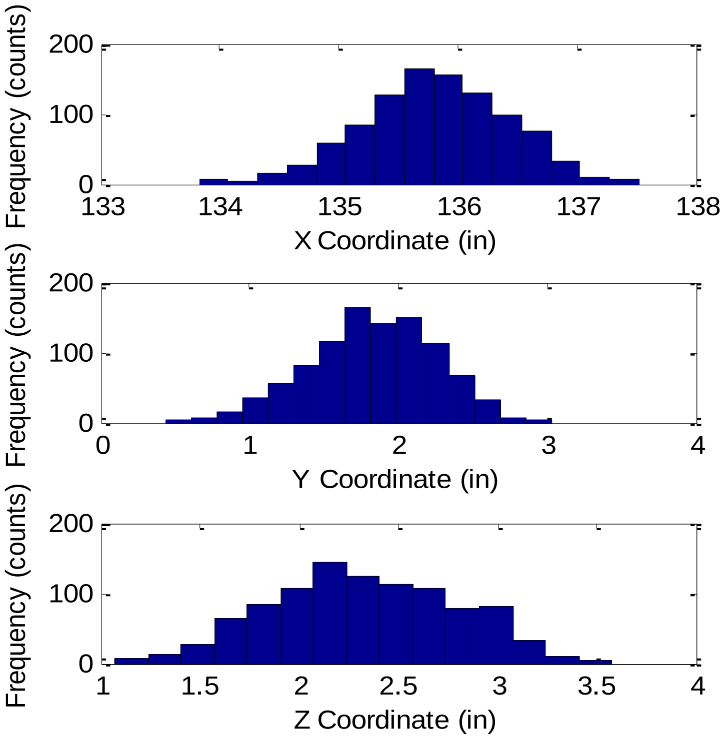
- The center of gravity calculation was solved using a numerical minimization algorithm
- Objective was to find the point that minimized the distance between the selected gravity vectors
- The minimum distance between a point and a line formed by two points is

$$d = \frac{|(\boldsymbol{x}_2 - \boldsymbol{x}_1) \times (\boldsymbol{x}_1 - \boldsymbol{x}_0)|}{|(\boldsymbol{x}_2 - \boldsymbol{x}_1)|}$$



- The minimum distance was found with respect to each gravity vector.
- The numerical algorithm minimized the sum of squares distances
- A Monte Carlo was run that perturbed the mean gravity vector directions by the standard deviations of each line

- The following results were obtained



	X (in)	Y (in)	Z (in)
Mean	135.81	1.83	2.33
2σ	1.23	0.87	0.93

Conclusion

- Four examples of measurement uncertainty and characterization projects were presented that highlighted the benefits of design of Experiments

Special Thanks to:

- In-Flight Force Measurement
  - Curt Larsen, NESC Johnson
- Non-monolithic Calibration Designs
  - NASA Aeronautics Test Program (ATP), GSRP program
- Variable Acceleration Calibration System
  - NASA LaRC Engineering Directorate, Co-op Program
- Center of Gravity Determination Method
  - Tom Jones, NASA LaRC
- Ray Rhew, Peter Parker and Drew Landman

Canonical Correlation Analysis for Longitudinal Data

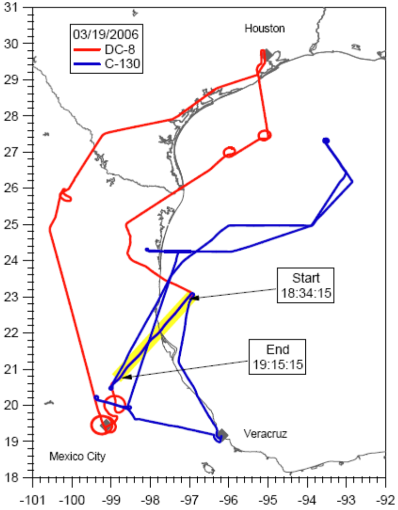
Raymond McCollum
Advisor Dayanand Naik

May 5th, 2010

INTEX

- The Intercontinental Chemical Transport Experiment (INTEX)
- "INTEX (http://cloud1.arc.nasa.gov) is a two phase experiment that aims to understand the transport and transformation of gases and aerosols on transcontinental/intercontinental scales and assess their impact on air quality and climate."
- The experiment was performed in the spring of 2006.
- The purpose of the project was to "Quantify the outflow and evolution of gases and aerosols from the Mexico City Megaplex".

Analysis Air Tracks*



INTEX*

- Multiple air frames will measure air and pollutants along the Mexican Coast.
- NASA DC-8 flown out o Houston, Texas
- NSF/NCAR C-130 from Tampico, Mexico
- Air frames will travel in close proximity.
- Data from multiple gasses will be recorded for each plane and compared in an effort to calibrate the instrumentation.

Canonical Correlation

Canonical correlation analysis (CCA) is used to identify and characterize the relationship between two sets of random vectors. Let Σ_x be the variance-covariance matrix of the $px1$ vector X, and let Σ_y be the variance covariance matrix of $qx1$ vector Y. Let the multivariate vectors X and Y have the covariance matrix Σ_{xy}.

$$\begin{bmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{bmatrix} \tag{1}$$

Classical Canonical Correlation*

The objective of canonical correlation analysis is to create a relationship between the X variables and the Y variables. CCA attempts to find a $qx1$ vector "a" and a $px1$ vector "b" to help define the correlation between the two data sets. The vectors a and b are chosen to maximize the correlation between $a'Y$ and $b'X$.

Repeated Canonical Correlation*

CCA was generalized to more than two sets of variables by Kettenring (1971)and other generalizations can be found in the literature. Let X and Y be repeated over time. Let $\mathbf{x}_i$ and $\mathbf{y}_i$ be vectors observed at the i^{th} time period. Represent the time periods by,

$$Y = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_t') \text{ and } X = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_t'). \tag{2}$$

Repeated Canonical Correlation*

For a CCA the with no additional time requirements, the covariance matrix is,

$$\begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix}. \tag{3}$$

Repeated Canonical Correlation*

The total number of parameters that require estimation becomes computationally intensive.

- The Σ_y matrix has $q(q+1)/2$ unique parameters.
- The Σ_x matrix has $p(p+1)/2$ unique parameters.
- The Σ_{xy} matrix has pq unique parameters for the cross correlations.

There are a total of $p(p+1)/2 + q(q+1)/2 + pq$ unique parameters that must be estimated. These values correspond to one set of parameters recorded at one time period. When sets of variables are recorded over time, the number of parameters required increases quickly.

Repeated Canonical Correlation*

Canonical Correlation Analysis for Longitudinal Data

Raymond McCollum
Advisor
Dayanand Naik

For t time periods the matrix becomes,

$$\begin{pmatrix} \Sigma_{y1y1} & \cdots & \Sigma_{y1yt} & \Sigma_{y1x1} & \cdots & \Sigma_{y1xt} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{yty1} & \cdots & \Sigma_{ytyt} & \Sigma_{ytx1} & \cdots & \Sigma_{ytxt} \\ \Sigma_{x1y1} & \cdots & \Sigma_{x1yt} & \Sigma_{x1x1} & \cdots & \Sigma_{x1xt} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{xty1} & \cdots & \Sigma_{xtyt} & \Sigma_{xtx1} & \cdots & \Sigma_{xtxt} \end{pmatrix} \tag{4}$$

This has a total of $\frac{t(p+q)(t(p+q)+1)}{2}$ parameters.

Repeated Canonical Correlation*

The Kronecker product covariance structure can be used to reduce the number of parameters. The variance covariance matrix of Y and X can be represented by the matrix below.

$$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix} \tag{5}$$

This matrix has considerably less parameters, namely

$$\frac{q(q+1) + p(p+1) + 2pq + 3t(t+1)}{2}. \tag{6}$$

Existing Solution*

Suppose $z_1, ..., z_N$ is a random sample of size N from the above multivariate normal distribution in (**??**). SNV (2008) obtained the maximum likelihood estimates of Σ and Ψ as follows. The MLE of an unrestricted positive definite matrix Σ is given as

$$\hat{\Sigma} = \frac{\sum_{i=1}^{N} z_{ic} \hat{\Psi}^{-1} z_{ic}'}{tN} \tag{7}$$

and similarly the MLE of an unrestricted matrix Ψ, except for the restriction that $\psi_{tt} = 1$, is given by

$$\hat{\Psi} = \frac{\sum_{i=1}^{N} z_{ic}' \hat{\Sigma}^{-1} z_{ic}}{pN} \tag{8}$$

and $\hat{\psi}_{tt} = 1$.

Existing Solution*

Here

$$z_i = \begin{pmatrix} z_{i11} & z_{i12} & \cdots & z_{i1t} \\ z_{i21} & z_{i22} & \cdots & z_{i2t} \\ \vdots & \vdots & \ddots & \vdots \\ z_{ip1} & z_{ip2} & \cdots & z_{ipt} \end{pmatrix},$$

$$z_{ic} = z_i - \bar{z}, \tag{9}$$

where,

$$\bar{z} = \begin{pmatrix} \bar{z}_{11} & \bar{z}_{12} & \cdots & \bar{z}_{1t} \\ \bar{z}_{21} & \bar{z}_{22} & \cdots & \bar{z}_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{z}_{p1} & \bar{z}_{p2} & \cdots & \bar{z}_{pt} \end{pmatrix}.$$

Existing Solution*

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1p} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2p} \\ \vdots & \ddots & & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pp} \end{pmatrix}$$

The solution to the Ψ multiplies the z_{ic} matrix by the Σ inverse estimate,

$$\hat{\Psi} = \frac{\sum_{i=1}^{N} z_{ic}' \hat{\Sigma}^{-1} z_{ic}}{Np},$$

Cross Correlation*

$$D = \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}. \tag{10}$$

Cross Correlation*

Estimates are,

$$\hat{\Psi}_{xy} = \frac{\sum_{i=1}^{N} X_{ic}' \hat{\Sigma}_{xy}^{+} Y_{ic}}{Rank(\Sigma_{xy})N} \tag{11}$$

$$\hat{\Sigma}_{xy} = \frac{\sum_{i=1}^{N} Y_{ic}' \hat{\Psi}_{xy}^{-1} X_{ic}}{Rank(\Psi_{xy})N}. \tag{12}$$

Partitioning Σ^*

Partitioning $\boldsymbol{\Sigma}$ will give

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}.$$

This can be broken up to get

$$\mathbf{D} = \begin{bmatrix} \Psi_{yy} \otimes \Sigma_{yy} & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_{xx} \otimes \Sigma_{xx} \end{bmatrix}.$$

Transformation*

Transform the data to make the diagonals of the covariance matrix the identity matrix.

Let

$$\begin{bmatrix} (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} & 0 \\ 0' & (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix} \tag{13}$$

Transformation*

The normal distribution was used to approximate the asymptotic distribution. Similar to Tan (1973) work.

$$\frac{approximately}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} I & D \\ D' & I \end{pmatrix}\right] \tag{14}$$

$$D = (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} C (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} \tag{15}$$

Transformation*

$$\begin{bmatrix} I & D \\ D' & I \end{bmatrix}^{-1/2} \begin{bmatrix} (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} & 0 \\ 0' & (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix} \tag{16}$$

$$\underset{\sim}{\underline{approximately}} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \right] \tag{17}$$

$$\begin{bmatrix} 0 & D \\ D' & 0 \end{bmatrix} \begin{bmatrix} I & D \\ D' & I \end{bmatrix}^{-1/2} \begin{bmatrix} (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} & 0 \\ 0' & (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} \end{bmatrix} \tag{18}$$

$$* \begin{bmatrix} Y \\ X \end{bmatrix} \underset{\sim}{\underline{approximately}} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} DD' & 0 \\ 0 & D'D \end{pmatrix} \right] \tag{19}$$

Unrestricted Covariance Matrix Ψ^*

The solution gives,

$$\widehat{D_{\Psi_{xy}} D'_{\Psi_{xy}}} \otimes \widehat{D_{\Sigma_{xy}} D'_{\Sigma_{xy}}}. \tag{20}$$

Spectral decomposition gives,

$$\widehat{D_{\Psi_{xy}} D'_{\Psi_{xy}}} = U \Delta^2 U' \Rightarrow \tag{21}$$

For $D_{\Psi_{xy}}$ positive definite we have the unique matrix (Harville (1997)).

$$\hat{D}_{\Psi_{xy}} = U \Delta U' \tag{22}$$

Unrestricted Covariance Matrix Ψ^*

$$C_{Base_{\Psi_{xy}}} = A_{\Psi_y}^{1/2} \hat{D}_{\Psi_{xy}} B_{\Psi_y}^{1/2} \tag{23}$$

Note that $C_{Base_{\Psi_{xy}}}$ can be a correlation matrix,

$$\hat{C}_{\Psi_{xy}} = Diag(\hat{C}_{Base_{\Psi_{xy}}})^{-1/2} C_{Base_{\Psi_{xy}}} Diag(\hat{C}_{Base_{\Psi_{xy}}})^{-1/2} \tag{24}$$

or $C_{Base_{\Psi_{xy}}}$ can use the AR(1) structure as noted earlier.

At this point we have a complete estimate for $\hat{C}_{\Psi_{xy}}$ to use to estimate $\hat{\Sigma}_{xy}$

$$\hat{\Sigma}_{xy} = \frac{\sum_{i=1}^{N} X_{ic} \hat{C}_{\Psi_{xy}}^{-1} Y'_{ic}}{NRank(C)} \tag{25}$$

Estimation Results*

Sample Size 500	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)$	$\|(\hat{\Theta} - \Theta)\|$
$\Sigma_y(11)$	5.50000	5.49091	0.19355	0.00909
$\Sigma_y(12)$	3.17543	3.16755	0.16785	0.00788
$\Sigma_y(13)$	6.50000	6.48315	0.23614	0.01685
$\Sigma_y(14)$	-0.50000	-0.50270	0.15944	0.00270
$\Sigma_y(21)$	3.17543	3.16691	0.19235	0.00852
$\Sigma_y(22)$	7.50000	7.48407	0.27176	0.01593
$\Sigma_y(23)$	0.35355	0.34927	0.14595	0.00428
$\Sigma_y(31)$	2.24537	2.23774	0.17092	0.00762
$\Sigma_y(32)$	-0.35355	-0.35161	0.17407	0.00195
$\Sigma_y(33)$	6.25000	6.23379	0.23151	0.01621
ρ_y	0.20000	0.19939	0.01497	0.00061
$\Sigma_x(11)$	4.55000	4.53781	0.18379	0.01219
$\Sigma_x(12)$	0.44907	0.44612	0.09253	0.00296
$\Sigma_x(21)$	2.70000	2.69447	0.10096	0.00553
ρ_x	0.40000	0.39907	0.01952	0.00093

Estimation Results*

Sample Size 500	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)$	$\lvert(\hat{\Theta} - \Theta)\rvert$
$\Sigma_{xy}(11)$	1.00416	1.00466	0.13027	0.00050
$\Sigma_{xy}(12)$	1.56525	1.56436	0.10695	0.00088
$\Sigma_{xy}(13)$	1.73925	1.73518	0.15310	0.00408
$\Sigma_{xy}(14)$	1.42009	1.41771	0.11413	0.00238
$\Sigma_{xy}(21)$	-0.27386	-0.27980	0.14816	0.00594
$\Sigma_{xy}(22)$	-0.22361	-0.22382	0.11304	0.00022
$\Sigma_{xy}(23)$	0.19365	0.18905	0.13287	0.00460
$\Sigma_{xy}(24)$	0.15811	0.15158	0.10581	0.00654
ρ_{xy}	0.30000	0.30217	0.04504	0.00217
$\Sigma 1_{xy}(11)$	1.00416	1.18419	0.45267	0.18003
$\Sigma 1_{xy}(12)$	1.56525	1.82970	0.61523	0.26446
$\Sigma 1_{xy}(13)$	1.73925	2.02502	0.67560	0.28576
$\Sigma 1_{xy}(14)$	1.42009	1.66430	0.57283	0.24420
$\Sigma 1_{xy}(21)$	-0.27386	-0.33653	0.21721	0.06267
$\Sigma 1_{xy}(22)$	-0.22361	-0.26029	0.16079	0.03669
$\Sigma 1_{xy}(23)$	0.19365	0.21089	0.16858	0.01724
$\Sigma 1_{xy}(24)$	0.15811	0.18202	0.14543	0.02391
$\rho_1 xy$	0.30000	0.34486	0.10577	0.04486

Estimation Results*

Sample Size 50	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)$	$\|(\hat{\Theta} - \Theta)\|$
$\Sigma_y(11)$	5.50000	5.38572	0.63838	0.11428
$\Sigma_y(12)$	3.17543	3.12812	0.53214	0.04730
$\Sigma_y(13)$	6.50000	6.37219	0.75108	0.12781
$\Sigma_y(14)$	-0.50000	-0.48519	0.52173	0.01481
$\Sigma_y(21)$	3.17543	3.10831	0.63379	0.06712
$\Sigma_y(22)$	7.50000	7.35267	0.87776	0.14734
$\Sigma_y(23)$	0.35355	0.35696	0.46468	0.00340
$\Sigma_y(31)$	2.24537	2.20256	0.55177	0.04281
$\Sigma_y(32)$	-0.35355	-0.35449	0.55607	0.00094
$\Sigma_y(33)$	6.25000	6.12585	0.74061	0.12415
ρ_y	0.20000	0.19905	0.04838	0.00095
$\Sigma_x(11)$	4.55000	4.43166	0.56430	0.11834
$\Sigma_x(12)$	0.44907	0.44469	0.29620	0.00438
$\Sigma_x(21)$	2.70000	2.65309	0.32754	0.04691
ρ_x	0.40000	0.39930	0.06129	0.00070

Estimation Results*

Sample Size 50	Simulations 1000	Converge		
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)$	$\|(\hat{\Theta} - \Theta)\|$
$\Sigma_{xy}(11)$	1.00416	0.99012	0.40499	0.01404
$\Sigma_{xy}(12)$	1.56525	1.55870	0.34521	0.00655
$\Sigma_{xy}(13)$	1.73925	1.70115	0.46706	0.03810
$\Sigma_{xy}(14)$	1.42009	1.41299	0.36314	0.00710
$\Sigma_{xy}(21)$	-0.27386	-0.27898	0.48118	0.00512
$\Sigma_{xy}(22)$	-0.22361	-0.21468	0.36631	0.00893
$\Sigma_{xy}(23)$	0.19365	0.19743	0.43212	0.00378
$\Sigma_{xy}(24)$	0.15811	0.17390	0.34203	0.01579
ρ_{xy}	0.30000	0.30405	0.07828	0.00405
$\Sigma1_{xy}(11)$	1.00416	2.21826	4.54869	1.21410
$\Sigma1_{xy}(12)$	1.56525	4.25959	10.03225	2.69434
$\Sigma1_{xy}(13)$	1.73925	4.30226	9.20887	2.56301
$\Sigma1_{xy}(14)$	1.42009	3.71180	8.74069	2.29170
$\Sigma1_{xy}(21)$	-0.27386	-0.67614	4.06503	0.40228
$\Sigma1_{xy}(22)$	-0.22361	-0.68435	4.28020	0.46074
$\Sigma1_{xy}(23)$	0.19365	0.63663	3.26688	0.44298
$\Sigma1_{xy}(24)$	0.15811	0.63454	3.27403	0.47643
$\rho1_{xy}$	0.30000	0.46671	0.35035	0.16671

Model II*

Model II allows a different time element correlation for each partition. That is, the X values have their own time correlation and the Y values have their own time correlation. The XY cross correlation value have their own separate time correlation. For example the time correlation for the X values at time one and the X values at time two may be .1, the Y values time one to time two correlation may be .3. The X at time one and Y at time two may have a .2 correlation.

$$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix} \tag{26}$$

Model III*

The model can retain the correlation structure in the time component of the Y values or the X values but not both. The cross correlation time structure retains its time correlation structure.

$$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & I_x \otimes \Sigma_x \end{pmatrix} \tag{27}$$

or

$$\begin{pmatrix} I_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix} \tag{28}$$

Either the Y or the X variable can be correlated in time. Hence the two equations 27 and 28 both have the same number of parameters to estimate.

$$\text{Parameters} = \frac{p(p+1)}{2} + \frac{q(q+1)}{2} + pq + 2 \tag{29}$$

Model IV*

This equation assumes the time correlation component is constant across all data. The mechanisms that are occurring in time play the same role in the X values, the Y values, and the cross project of the two. The corresponding matrix is shown in equation 30.

$$\begin{pmatrix} \Psi_t \otimes \Sigma_y & \Psi_t \otimes \Sigma_{yx} \\ \Psi_t \otimes \Sigma_{xy} & \Psi_t \otimes \Sigma_x \end{pmatrix} \tag{30}$$

In this model, all ρ parameters are equal.

$$\rho_x = \rho_y = \rho_{xy} \tag{31}$$

$$\text{Parameters} = \frac{p(p+1)}{2} + \frac{q(q+1)}{2} + pq + 1 \tag{32}$$

Model V*

The time correlation matrix Ψ equals the identity matrix throughout all four partitions.

$$\begin{pmatrix} I_y \otimes \Sigma_y & I_{yx} \otimes \Sigma_{yx} \\ I_{xy} \otimes \Sigma_{xy} & I_x \otimes \Sigma_x \end{pmatrix} \tag{33}$$

Model 33 shows no covariance structure across time units. This model assumes that what happens in time unit 1 does not influence time unit 2 or later. This simple analysis may be what a researcher will attempt when first faced with multivariate time series data.

Model 33 has the least number of parameters that require estimation.

$$\text{Model 33 Parameters} = \frac{p(p+1)}{2} + \frac{q(q+1)}{2} + pq \tag{34}$$

Hypothesis Testing

- Model I : completely unstructured covariance
- Model II : Kronecker product covariance structure with DC-8 and C-130 each having a different time correlation
- Model III : Kronecker product covariance structure where DC-8 or C-130 has no time correlation
- Model IV : Kronecker product covariance where DC-8 and C-130 have the same time correlation (A true maximum likelihood solution exists under transformation)
- Model V : Kronecker product covariance where there is no time correlation (A true maximum likelihood solution exists under transformation)

Hypothesis Testing*

The log likelihood ratio test was used to determine if the model should be increased in complexity. Note that the data is not normally distributed but this model was used as an approximate test statistics.

$$f(Z,\mu,\Sigma) = \frac{1}{2\pi^{(p+q)t/2}|\Sigma|^{1/2}} exp \frac{-(\mathbf{Z}')\Sigma^{-1}(\mathbf{Z})}{2} \tag{35}$$

$$\mathbf{Z} = \begin{bmatrix} Y \\ X \end{bmatrix} \tag{36}$$

The log likelihood ratio test was used as a test statistic.

$$-2(log f(Z,0,\Sigma_{H_o}) - log f(Z,0,\Sigma_{H_a})) \sim \chi^2_{df} \tag{37}$$

Hypothesis Testing*

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
II	500	.115	1.000	1.000	.782	1.000
II	350	.097	1.000	1.000	.290	1.000
II	200	.147	1.000	.999	.011	1.000
II	100	*	1.000	.993	0	1.000
II	50	*	.975	.982	0	1.000
III	500	.137	.109	1.000	1.000	1.000
III	350	.114	.105	1.000	1.000	1.000
III	200	.157	.094	1.000	.994	1.000
III	100	*	.084	1.000	.340	.965
III	50	*	.063	.996	.005	.732

Table: Rejection rates for 1000 samples

Hypothesis Testing*

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
IV	500	.113	1.000	.917	0	1.000
IV	350	.145	1.000	.923	0	1.000
IV	200	.145	1.000	.928	0	1.000
IV	100	*	1000	.889	0	1.000
IV	50	*	1000	.775	0	1.000
V	500	.122	.099	.021	0	.032
V	350	.091	.089	.015	0	.018
V	200	.142	.095	.020	0	.053
V	100	*	.080	.015	0	.038
V	50	*	.051	.023	0	.027

Table: Rejection rates for 1000 samples

Bootstrapping*

The hypothesis tests above showed a higher rejection rate than the expected .05. This was most likely due to differences between the MLE and the estimates. Bootstrapping was used to create tests that give more accurate rejection probabilities. Parametric bootstrapping stimulations based on (Efron and Tibshirani(1993)) theory were used to create hypothesis tests. For each possible covariance structure, a set of 100 simulations of 100 bootstrap samples each were used.

Bootstrap samples*

Initial data set.

$$\begin{pmatrix} y_{111} & \cdots & y_{p11} & \cdots & y_{1t1} & \cdots & y_{pt1} & x_{111} & \cdots & x_{q11} & \cdots & x_{1t1} & \cdots & x_{qt1} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ y_{11n} & \cdots & y_{p1n} & \cdots & y_{1tn} & \cdots & y_{ptn} & x_{11n} & \cdots & x_{q1n} & \cdots & x_{1tn} & \cdots & x_{qtn} \end{pmatrix}' \quad (38)$$

The likelihood ratio statistics was used as a test statistic for the bootstrap.

$$\lambda = -2(log(L(Y, X, \mu, \Sigma_{H_o})) - log(L(Y, X, \mu, \Sigma_{H_a}).)$$

Bootstrapping*

The initial data was used to generate a bootstrap estimate of the variance covariance matrix

$$\Phi = \begin{pmatrix} \hat{\Psi}_y \otimes \hat{\Sigma}_y & \hat{\Psi}_{yx} \otimes \hat{\Sigma}_{yx} \\ \hat{\Psi}_{xy} \otimes \hat{\Sigma}_{xy} & \hat{\Psi}_x \otimes \hat{\Sigma}_x \end{pmatrix}.$$

The bootstrap variance covariance matrix was used to generate B bootstrap samples.

$$\begin{gathered} y_1^*, y_2^*, \ldots, y_b^* \\ x_1^*, x_2^*, \ldots, x_b^* \end{gathered}$$

Bootstrapping*

Each bootstrap sample was used to create an estimate of the sample variance covariance matrix.

$$\Phi^* = \begin{pmatrix} \hat{\Psi}^*_y \otimes \hat{\Sigma}^*_y & \hat{\Psi}^*_{yx} \otimes \hat{\Sigma}^*_{yx} \\ \hat{\Psi}^*_{xy} \otimes \hat{\Sigma}^*_{xy} & \hat{\Psi}^*_x \otimes \hat{\Sigma}^*_x \end{pmatrix}.$$

Each bootstrap sample was also used to get a test statistics.

$$\lambda^*_b = -2(log(L(Y, X, \mu, \Sigma_{H^*_0})) - log(L(Y, X, \mu, \Sigma_{H^*_{(1)}}))).$$

λ was compared to the vector of λ^*_bs to get an estimate of the p-value.

Bootstrapping Hypothesis Tests*

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
II	500	.08	1.0	1.0	1.0	1.00
II	350	.07	1.0	1.0	.99	1.00
II	200	.06	1.0	1.0	.98	1.00
II	100	*	.92	1.0	.92	1.00
II	50	*	.95	.96	.66	1.00
III	500	.06	.05	1.00	1.00	1.00
III	350	.1	.06	1.00	1.00	1.00
III	200	.05	.01	1.00	1.00	1.00
III	100	*	.06	1.00	1.00	.96
III	50	*	.02	1.00	1.00	.77

Table: Rejection rates for 100 samples, Null Hypothesis II

Bootstrapping Hypothesis Tests*

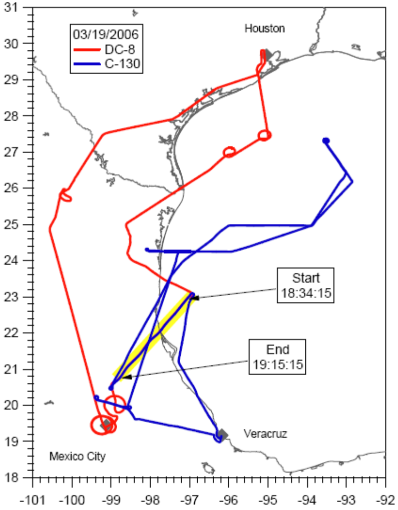
Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
IV	500	.05	1.00	.94	.04	1.0
IV	350	.1	1.00	.89	.07	1.0
IV	200	.03	1.00	.94	.06	1.0
IV	100	*	1.00	.88	.06	1.0
IV	50	*	1.00	.88	.04	1.0
V	500	.05	.08	.1	.1	.07
V	350	.10	.08	.03	.07	.07
V	200	.03	.08	.09	.11	.06
V	100	*	.05	.08	.1	.04
V	50	*	.05	.05	.06	.06

Table: Rejection rates for 100 samples, Null Hypothesis II

INTEX*

- The Intercontinental Chemical Transport Experiment (INTEX)
- "INTEX (http://cloud1.arc.nasa.gov) is a two phase experiment that aims to understand the transport and transformation of gases and aerosols on transcontinental/intercontinental scales and assess their impact on air quality and climate."
- The experiment was performed in the spring of 2006.
- The purpose of the project was to "Quantify the outflow and evolution of gases and aerosols from the Mexico City Megaplex".

Analysis Air Tracks*



INTEX*

- Multiple air frames will measure air and pollutants along the Mexican Coast.
- NASA DC-8 flown out o Houston, Texas
- NSF/NCAR C-130 from Tampico, Mexico
- Air frames will travel in close proximity.
- Data from multiple gasses will be recorded for each plane and compared in an effort to calibrate the instrumentation.

INTEX*

- Three pollutants of interest for this study were H2O Water CO Carbon Monoxide O3 Ozone.
- Data were recorded over time and the three gasses were thought to be correlated.
- Sensors give different readings and neither sensor is considered to be the "correct" answer.
- The objective is to study the covariance structure of sensor measurement on both airframes. The structure will reveal how each plane's sensor readings vary with time.

Altitude and Molecule Measurements*

Altitude	O3 DC-8	CO DC-8	H20 DC-8	O3 C-130	CO C-130	H2O C-130
313	35.0317	112.93	2.295522603	33.2	103.3845	3.90521
3992.3	81.85337	222.59	14.55818318	85.2	211.3836	16.5065

Table: Altitude and Molecule Measurements before scaling

Units for CO and O3 are ppbv: Number of molecules per cubic centimeter over number of air molecules per cubic centimeter. Units for H20 are g/kg: grams of water vapor per kg dry air.

Altitude and Molecule Measurements*

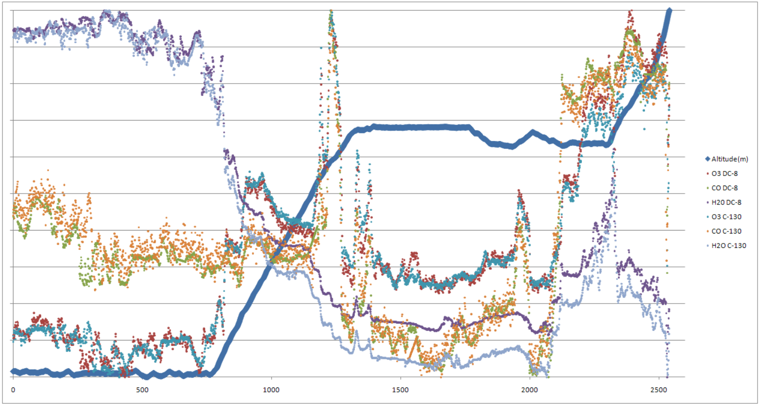


Figure: INTEX-B Airtracks Altitude

Data Hypothesis Test Results*

Hypothesis Test	Observed P-value
II $\Longrightarrow$ III	.00
II $\Longrightarrow$ IV	.23
IV $\Longrightarrow$ V	.007

Table: Testing Results for INTEX data

Variance Covariance Matrices*

Σ_yDC-8 O3	Σ_yDC-8 CO	Σ_yDC-8 H20
0.2119047	-0.011894	0.0049709
-0.011894	0.5828666	-0.01056
0.0049709	-0.01056	0.0053191

Table: Covariance IV Estimates for NASA Σ_y

Σ_xC-130 O3	Σ_x C-130 CO	Σ_xC-130 H20
0.2359842	0.1215785	0.0014622
0.1215785	7.2159748	0.0053734
0.0014622	0.0053734	0.011764

Table: Covariance IV Estimates for NASA Σ_x

Variance Covariance Matrices*

Σ_{xy} O3	Σ_{xy} CO	Σ_{xy} H20
0.0371183	-0.056899	-0.00096
0.0284117	-0.0285	0.002261
-0.000511	0.0094509	6.8809E-6

Table: Covariance IV Estimates for NASA Σ_{xy}

Ψ O3	Ψ CO	Ψ H20
1	-0.131819	0.0173763
-0.131819	1	-0.131819
0.0173763	-0.131819	1

Table: Covariance IV Estimates for NASA Ψ

Canonical Correlations*

Canonical Correlation	Correlation	Cumulative Percentage
1*st* Canonical Correlation	0.201539	.726285
2*nd* Canonical Correlation	0.046524	.893943
3*rd* Canonical Correlation	0.02943	1

Table: Canonical Correlations Within Each Time Period

Canonical Coefficients*

	1st Variable	2nd Variable	3rd Variable
Ozone DC-8	0.9296	-0.1022	-0.3845
Carbon M. DC-8	0.3716	0.6197	0.7178
Water DC- 8	-0.2272	0.9193	-0.4032

Table: Standardized Canonical Coefficients DC-8

	1st Variable	2nd Variable	3rd Variable
Ozone C-130	0.9602	0.2816	-0.0899
Carbon M. C-130	-0.3806	0.8357	-0.4072
Water C-130	-0.0589	0.3999	0.9152

Table: Standardized Canonical Coefficients C-130

Conclusion*

- Repeated CCA is a method that allows the comparison of multiple random variables to each other.
- The procedure is distribution independent and estimates the the variance covariance.
- The number of variables required to estimate variance covariance matricies grows quickly.
- Modeling the data struction in accordance with subject matter expert knowledge reduces the data requirements.

- Bakewell, D., and Wit, E., "Bartlett Correction for likelihood ratio test", *Unpublished report*, 3 pages.
- Bradley, E. and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.
- Chaganty, N. R., and Naik, D. N. (2002), "Analysis of Multivariate Longitudinal Data using Quasi-Least Squares", *Journal of Statistical Planning and Inference*, 103, 421-436.
- Harville, D. A. (1973), *Matrix Alegebra From a Statistician's Perspective*, New York: Springer.
- Hotelling, H. (1936), "Relations Between Two Sets of Variates", *Biometrika*, 28, 321-377.

- Johnson, R. A., and Wichern, D. W., (2002), *Applied Multivariate Statistical Analysis*, New Jersey: Prentice-Hall.
- Kettenring, J. R. (1971), "Canonical Analysis of Several Sets of Variables", *Biometrika*, 58, 433-451.
- Khattree, R., and Naik, D. N. (2000), *Multivariate Data Reduction and Discrimination with SAS Software.*, North Carolina: Wiley-SAS.
- Mardia, K. V., Kent, J. J.(1979), & Bibby, J. M. (1979), *Multivariate Analysis*, New York: Academic Press.
- Naik, D. N., and Rao, S. (2001), "Analysis of Multivariate Repeated Measures Data with a Kronecker Product Structured Covariance Matrix", *Journal of Applied Statistics*, 28, 91-105.

Ravishanker, N. and Dey, D (2002), *A First Course in Linear Models Theory*, New York: Chapman & Hall/CRC.

Roy, A., and Khattree, R. (2005), "On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data", *Statistical Methodology*, 2, 297-306.

Schott, J. R. (1997), *Matrix Analysis for Statistics*, New York: John Wiley & Sons, Inc.

Srivastava J. (2007), "Canonical Variate Analysis and Related Methods with Longitudinal Data", *Ph.D. thesis, Department of Statistics, Old Dominions University*.

Srivastava, M. S., Nahtman, T., and von Rosen, D. (2008), "Estimation in general multivariate linear models with

Kronecker product covariance structure", *Research Report, Swedish University of Agricultural Sciences*, 21 pages.

- Srivastava, M.S., Nahtman, T., and von Rosen, D.(2008), "Models with a Kronecker Product Covariance Structure: Estimation and Testing", *Mathematical Methods of Statistics*, 17(4), 357-370.
- Srivastava, J and Naik, D. N. (2008), "Canonical Correlation Analysis of Longitudinal Data", *Denver JSM 2008 Proceedings, Biometrics Section*, 563-568.
- Tan, W. Y. (1973), "Multivariate Studentization and its Applications", *The Canadian Journal of Statistics*, V1, N2, 181-199.

Questions?

Doing the Right Things: Leaders Wanted … Apply Within Sounding the Call to Arms

Greg Hutto & Jim Simpson
Ops Analysts Test Wings
Air Armament Center
Eglin AFB, Florida
Gregory.hutto@eglin.af.mil

NASA Statistical Engineering Symposium 5 May 2011

Structure

- If this DOE stuff is so good … why do I struggle?
- Outline of a story to convince our leaders
- Equipping leaders with the right questions to ask
- Summary & Questions

If all this DOE Stuff is so good … why do I struggle?

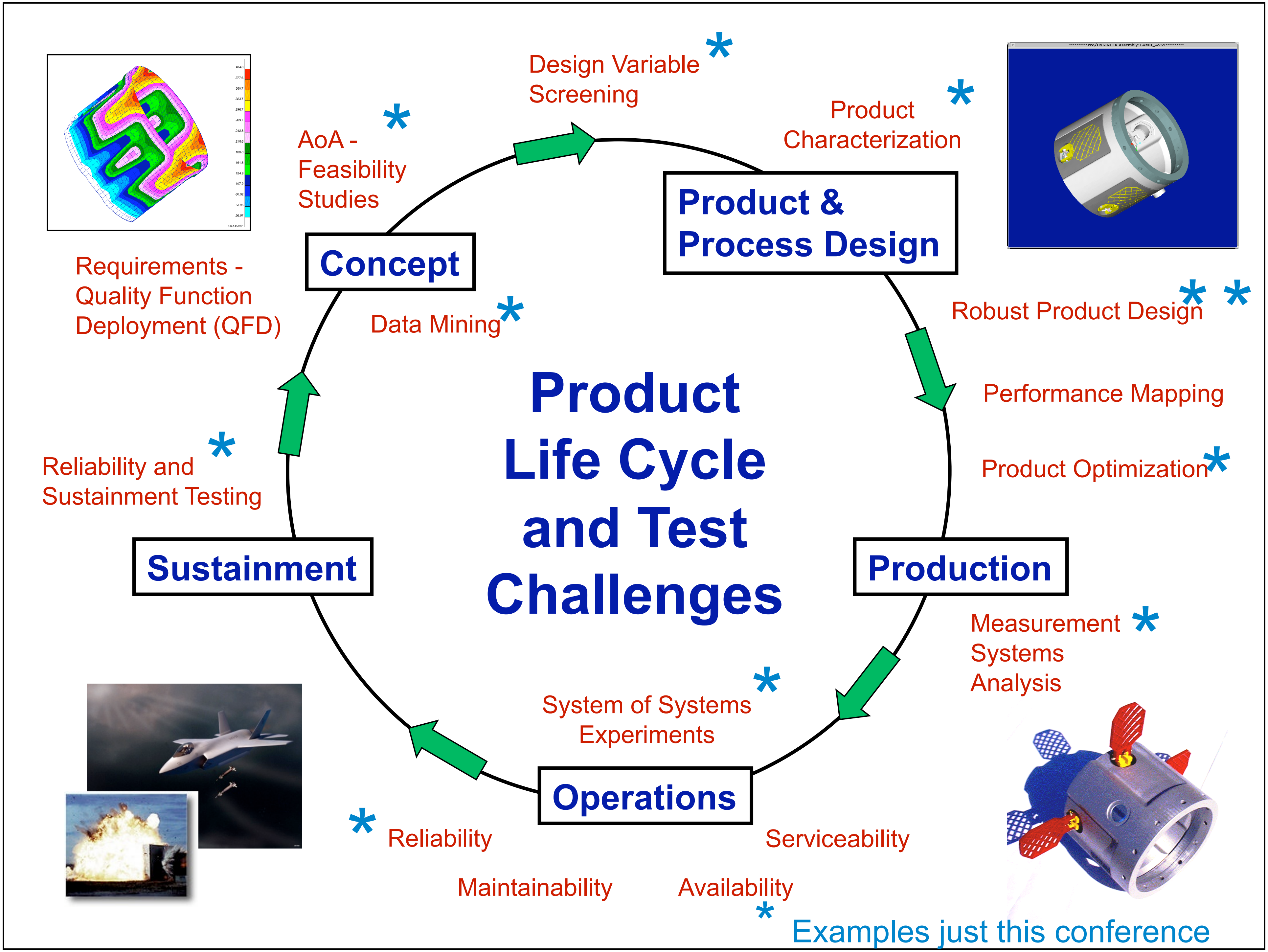
Deming and the VP – May be Apocryphal, but True …

"Learning is not compulsory . . . neither is survival.“

"It is not enough to do your best; you must know *what* to do, and then do your best."

-- W. Edwards Deming

October 14, 1900 – December 20, 1993

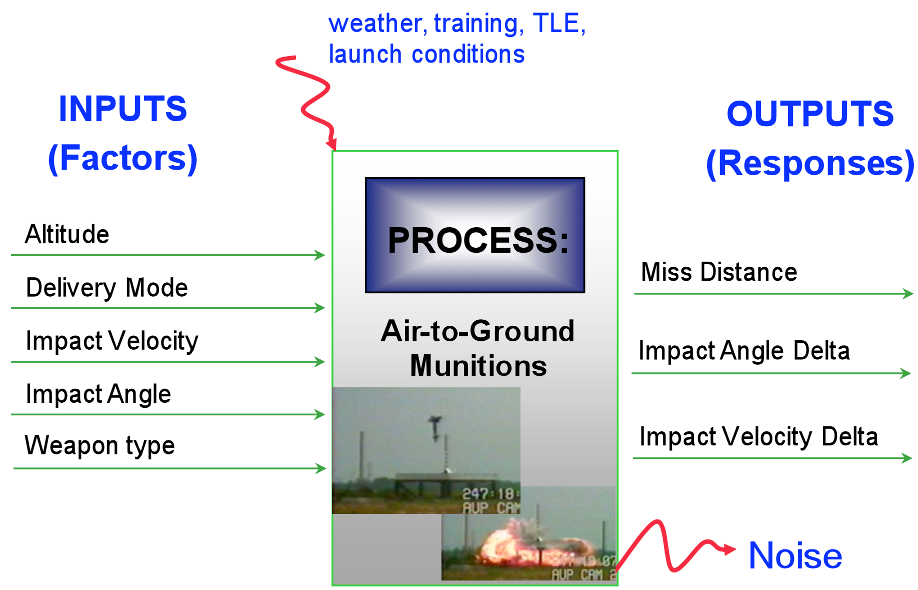

Product
Life Cycle
and Test
Challenges
Concept
Product &
Process Design
Production
Operations
Sustainment
Design Variable
Screening
AoA -
Feasibility
Studies
Product
Characterization
Requirements -
Quality Function
Deployment (QFD)
Data Mining
Robust Product Design
Performance Mapping
Product Optimization
Reliability and
Sustainment Testing
Measurement
Systems
Analysis
System of Systems
Experiments
Reliability
Serviceability
Maintainability
Availability
* Examples just this conference

Systems Engineering Employ Many Simulations of Reality

<table>
<tr><td></td><td></td><td colspan="8">Simulation of Reality</td></tr>
<tr><td colspan="2">Acq Phase</td><td colspan="2">M&S</td><td colspan="2">Hardware</td><td colspan="2">System/Flight Test</td></tr>
<tr><td>Reqt'ts Dev</td><td></td><td rowspan="7">Warfare</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>AoA</td><td rowspan="6">Physics</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Concepts</td><td></td><td rowspan="5">HWIL/SIL</td><td rowspan="5">Captive</td><td rowspan="5">Subsystem</td><td rowspan="2">Prototype</td></tr>
<tr><td></td><td>Risk Reduction</td></tr>
<tr><td>EMD</td><td></td><td rowspan="2">Prod Rep</td></tr>
<tr><td></td><td>Prod & Mfr</td></tr>
<tr><td>Sustain</td><td></td><td>Production</td></tr>
</table>

- At each stage of development, we conduct experiments
  - Ultimately – how will this device function in service (combat)?
  - Simulations of combat differ in fidelity and cost
  - Differing goals (screen, optimize, characterize, reduce variance, robust design, trouble-shoot)
  - Same problems – distinguish truth from fiction: What matters? What doesn't?

What are Statistically Designed Experiments?



- Purposeful, systematic changes in the inputs in order to observe corresponding changes in the outputs
- Results in a mathematical model that predicts system responses for specified factor settings

$$\text{Responses} = f\left(\text{Factors}\right) + \varepsilon$$

Case DT/OT: B-1 Radar TLE Accuracy Characterization (2001)

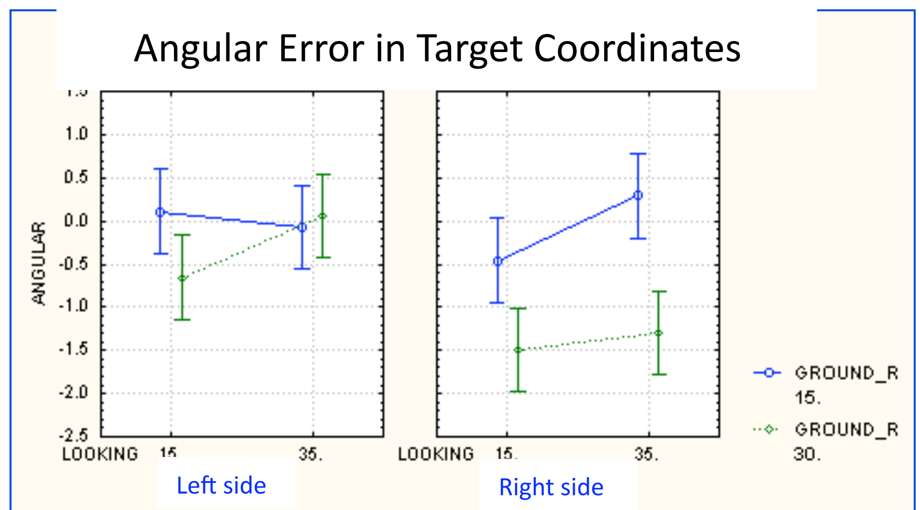
Problem:

- Is B-1B APQ-164 monopulse SAR mode for targeting accurate enough for JDAM?
- Are tail numbers similar? Target types?
- Bottom line: self-target JDAM?
- 7 sorties flown with mixed results
  -100's of measurements "as available"

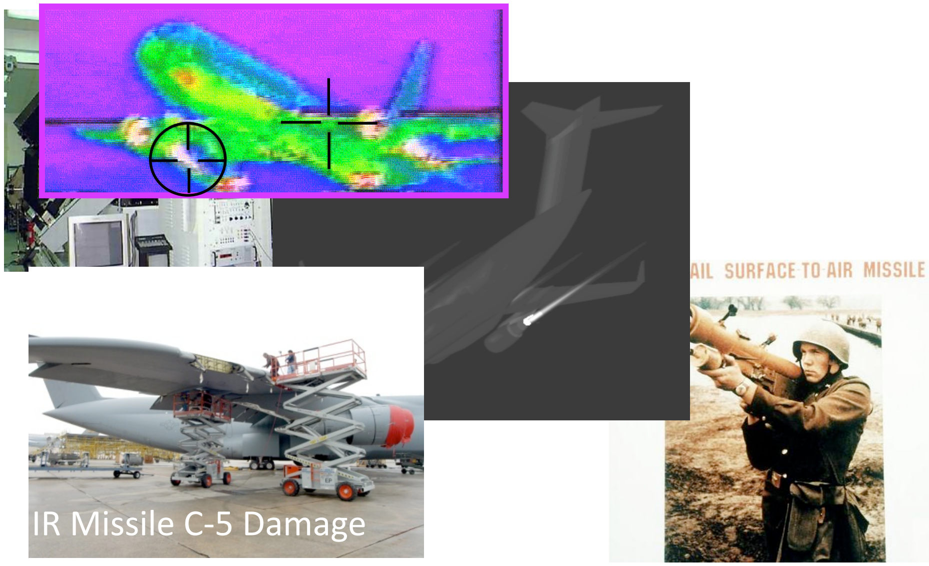
DOE Approach:

- Variables include
  - Side of A/C, angle off nose
  - Range, type of target
  - Two tail numbers
- Responses include TLE, mil error
- Compare to specified radar accuracy
- Single 2-ship sortie

Results: Similar accuracy across volume, tail



Case: DT HWIL GWEF Large Aircraft IR Hit Point Prediction



Test Objective:

- IR man-portable SAMs pose threat to large aircraft in current AOR
- Dept Homeland Security desired Hit point prediction for a range of threats needed to assess vulnerabilities
- Solution was HWIL study at GWEF (ongoing)

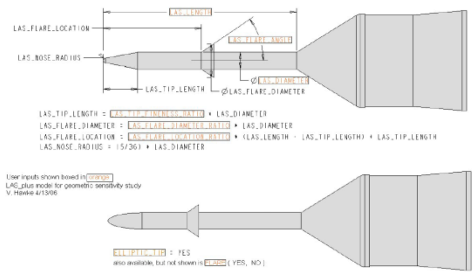
DOE Approach:

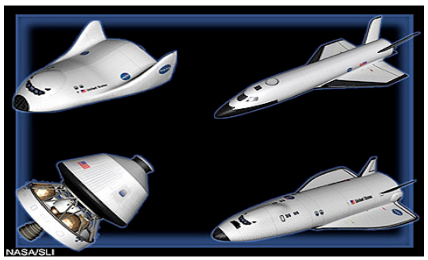
- Aspect – 0-180 degees, 7each
- Elevation – Lo,Mid,Hi, 3 each
- Profiles – Takeoff, Landing, 2 each
- Altitudes – 800, 1200, 2 each
- Including threat – 588 cases
- With usual reps nearly 10,000 runs
- DOE controls replication to min needed

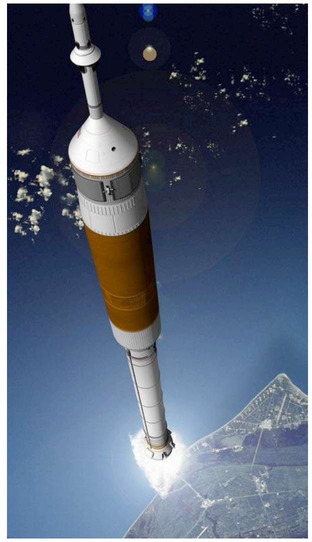
Results:

- Revealed unexpected hit point behavior
- Process highly interactive (rare 4-way)
- Process quite nonlinear w/ 3^{rd} order curves
- Reduced runs required 80% over past
- Possible reduction of another order of magnitude to 500-800 runs

Case 11: CFD for NASA CEV







Test Objective:

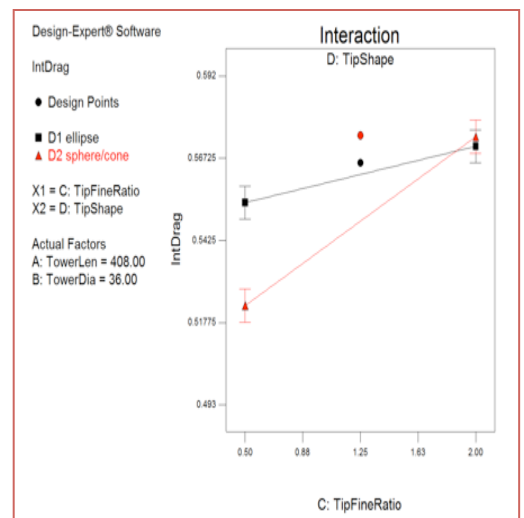
- Select geometries to minimize total drag in ascent to orbit for NASA's new Crew Exploration Vehicle (CEV)
- Experts identified 7 geometric factors to explore including nose shape
- Down-selected parameters further refined in following wind tunnel experiments

DOE Approach:

- Two designs – with 5 and 7 factors to vary
- Covered elliptic and conic nose to understand factor contributions
- Both designs were first order polynomials with ability to detect nonlinearities
- Designs also included additional confirmation points to confirm the empirical math model in the test envelope

Results:

- Original CFD study envisioned 1556 runs
- DOE optimized parameters in 84 runs – 95%!
- ID'd key interaction driving drag



Source: A Parametric Geometry CFD Study Utilizing DOE *Ray D. Rhew, Peter A. Parker, NASA Langley Research Center,* **AIAA 2007 1616**

So … *why* aren't *all* experiments well-designed?

- Summary of three projects:
  - 1 mission when 7 couldn't answer the question
  - Cut runs from 5000 replicates to 500
  - CFD Trials reduced from 1920 to 84
- Many such outstanding success stories
- We know how to teach & mentor practitioners
- Experts can be hired and groomed
- We have plenty of good software tools, texts

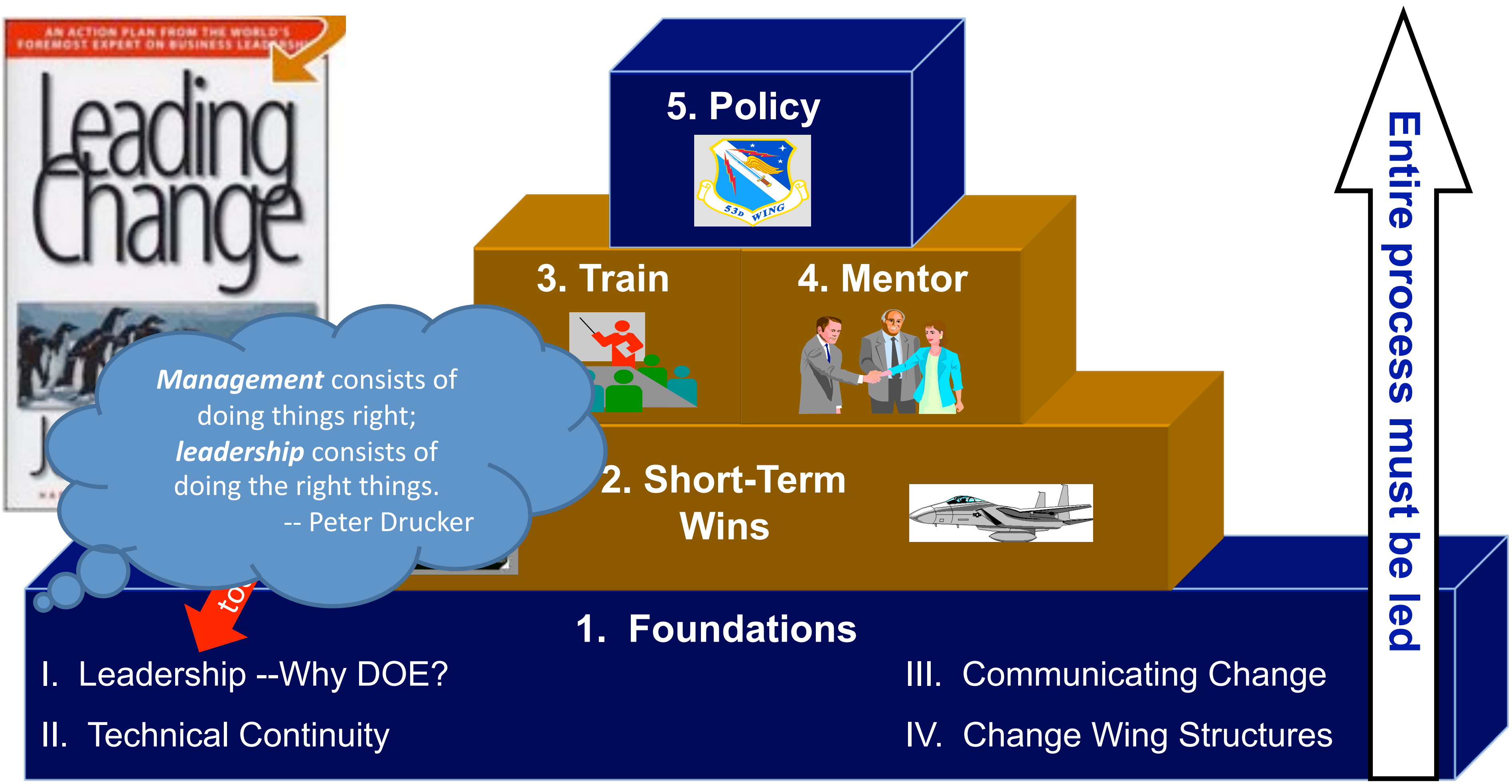
"We have met the enemy and he is … Us!
-- Pogo circa 1971

- It is us…
- A Job Story circa 1990-2000
- "Leadership From Below"

  -- Col T.S. Hutto 1933-1998

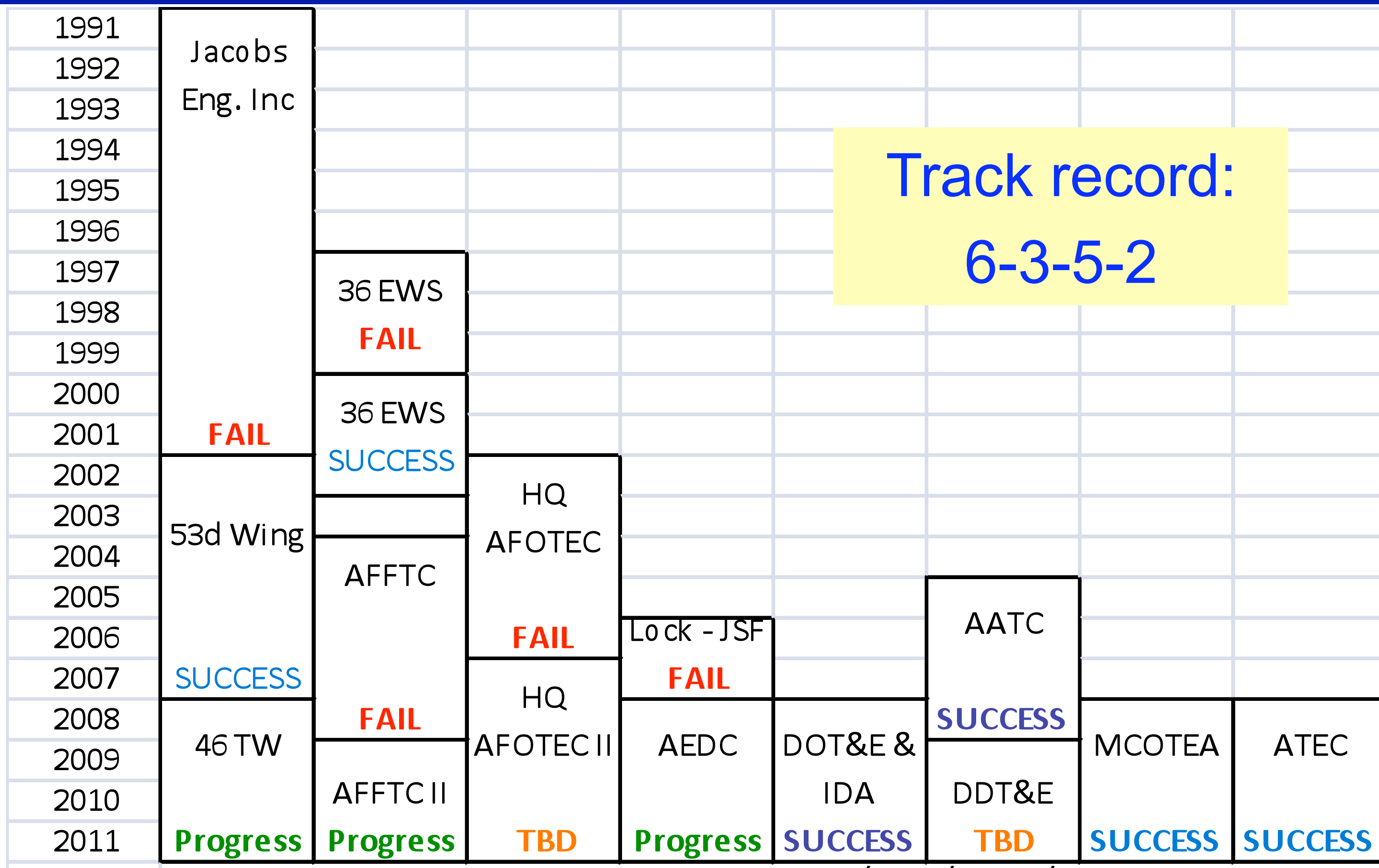
"But how can people call on him if they have not believed in him? How can they believe in him if they have not heard his message? How can they hear if no one tells the Good News? " -- Paul (0063, Romans 10.14)

Five Steps to Implementation



"Because **management** deals mostly with the **status quo** and **leadership** deals mostly with **change**, in the next century we are going to have to try to become much more skilled at creating leaders." -- Dr. John Kotter

Telling the "Why?" Story … It is not easy or guaranteed of success



FAIL = Pockets of success but exec not organize/train/equip/measure to sustain

PROGRESS = Efforts to organize/train/equip/hire and accountability by senior exec

TBD = Encouraging engagements with staff, executives

SUCCESS = Exec establishes accountability, resources, hires, policy. Majority DOE

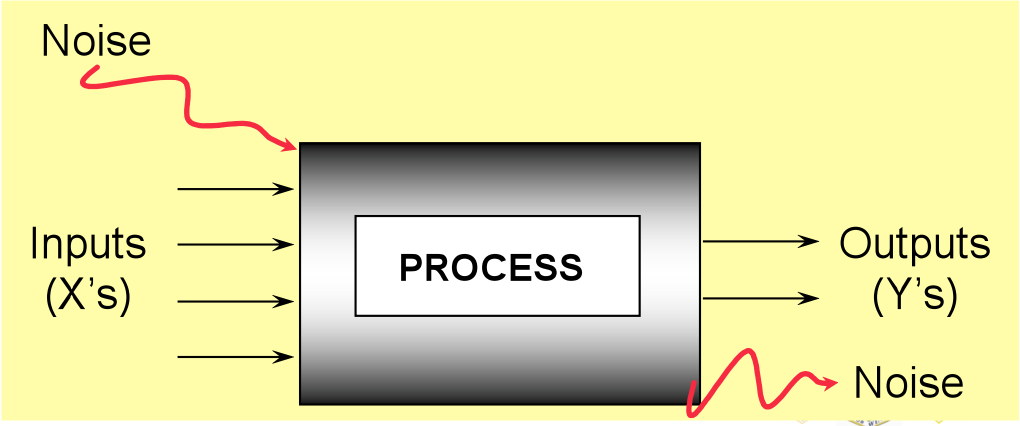
Why DOE? One Slide…
DOE Gives Scientific Answers to <u>Four</u> Fundamental Test Challenges

Four Challenges faced by any test

1. *How many? <u>Depth</u> of Test* – effect of test size on uncertainty
2. *Which Points? <u>Breadth</u> of Testing* – spanning the vast employment battlespace
3. *How Execute? <u>Order</u> of Testing* – insurance against "unknown-unknowns"
4. *What Conclusions? Test <u>Analysis</u>* – drawing objective, scientific conclusions while controlling noise

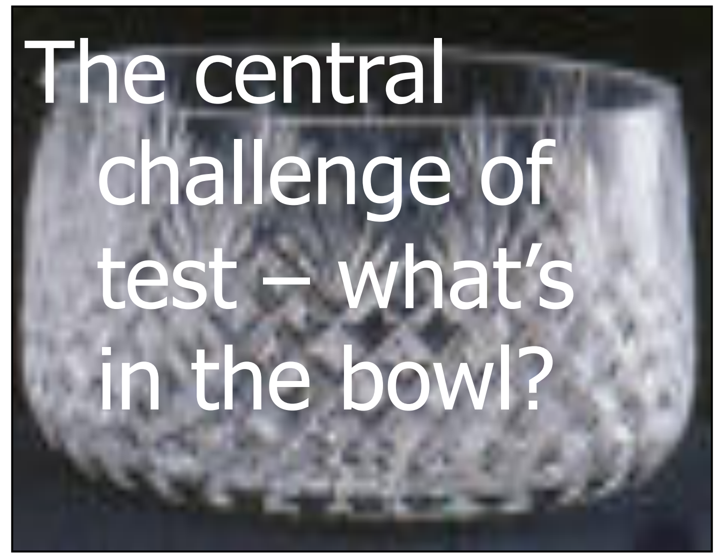
DOE effectively addresses all these challenges!

In our short time today, address primarily #1 and #2.

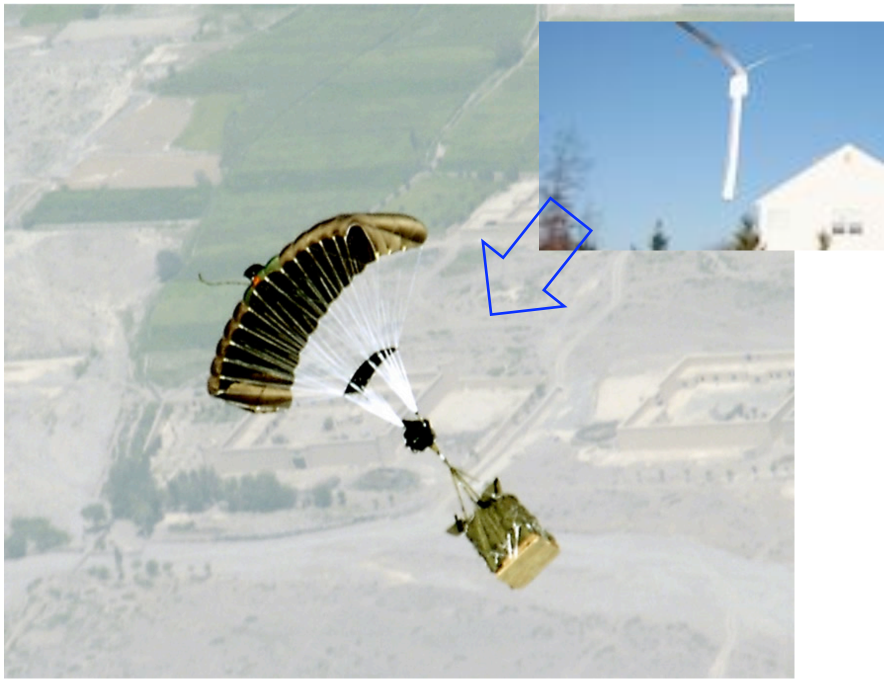


Question #1 … How Many?

- In all our testing – we reach into the bowl (reality) and draw a sample of JPADS performance
- Consider an "80% JPADS"
  - Suppose a required 80% P(Arrival)
  - Is the Concept version acceptable?
- We don't know in advance which bowl God hands us …
  - The one where the system *works* or,
  - The one where the system *doesn't*



Example:
Precision Air Drop System



The dilemma for airdropping supplies has always been a stark one. High-altitude airdrops often go badly astray and become useless or even counter-productive. Low-level paradrops face significant dangers from enemy fire, and reduce delivery range. Can this dilemma be broken?
A new advanced concept technology demonstration shows promise, and is being pursued by U.S. Joint Forces Command (USJFCOM), the U.S. Army Soldier Systems Center at Natick, the U.S. Air Force Air Mobility Command (USAF AMC), the U.S. Army Project Manager Force Sustainment and Support, and industry. The idea? Use the same GPS-guidance that enables precision strikes from JDAM bombs, coupled with software that acts as a flight control system for parachutes. JPADS (the Joint Precision Air-Drop System) has been combat-tested successfully in Iraq and Afghanistan, and appears to be moving beyond the test stage in the USA… and elsewhere.

Capability:

Assured SOF re-supply of material

Requirements:

Probability of Arrival
Unit Cost $XXXX
Damage to payload
Payload
Accuracy
Time on target
Reliability …

- Just when you think of a good class example – they are already building it!
- 46 TS – 46 TW Testing JPADS

Start -- Blank Sheet of Paper: How Many?

- Let's draw a sample of _*n*_ drops
- How many is enough to get it *right*?
  - 3 – because that's how much $/time we have
  - 8 – because I'm an 8-guy
  - 10 – because I'm challenged by fractions
  - 30 – because something good happens at 30!

- Let's start with 10 and see …

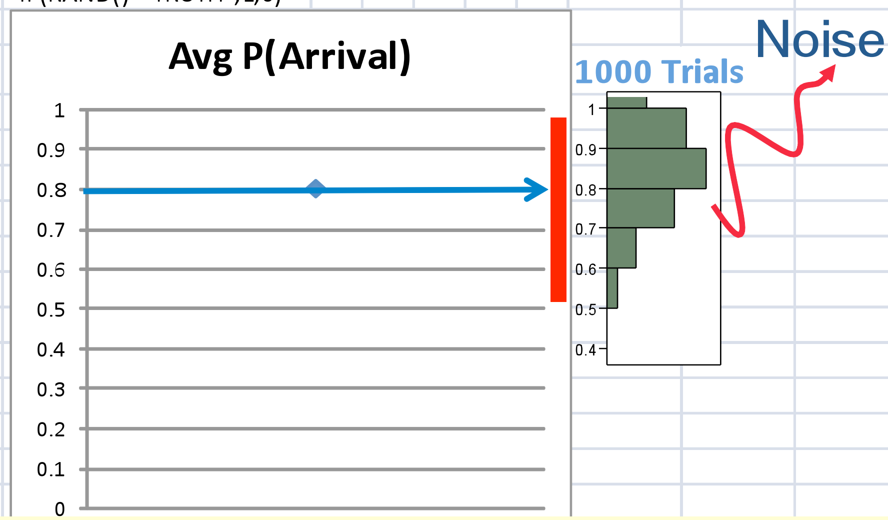
=> Switch to Excel File – JPADS Pancake.xls

Embedded Excel Simulation to Address "How Many?"

True P(Arrival)	80%
Trial/Rep	**Miss?**
1	1
2	1
3	1
4	1
5	1
6	1
7	0
8	1
9	0
10	1
AvgMiss	**0.8**

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

=IF(RAND()<"TRUTH",1,0)



Definitions:

α - false positive error rate - concluding a difference exists (good or bad) when the difference is noise.

Confidence is 1-α.

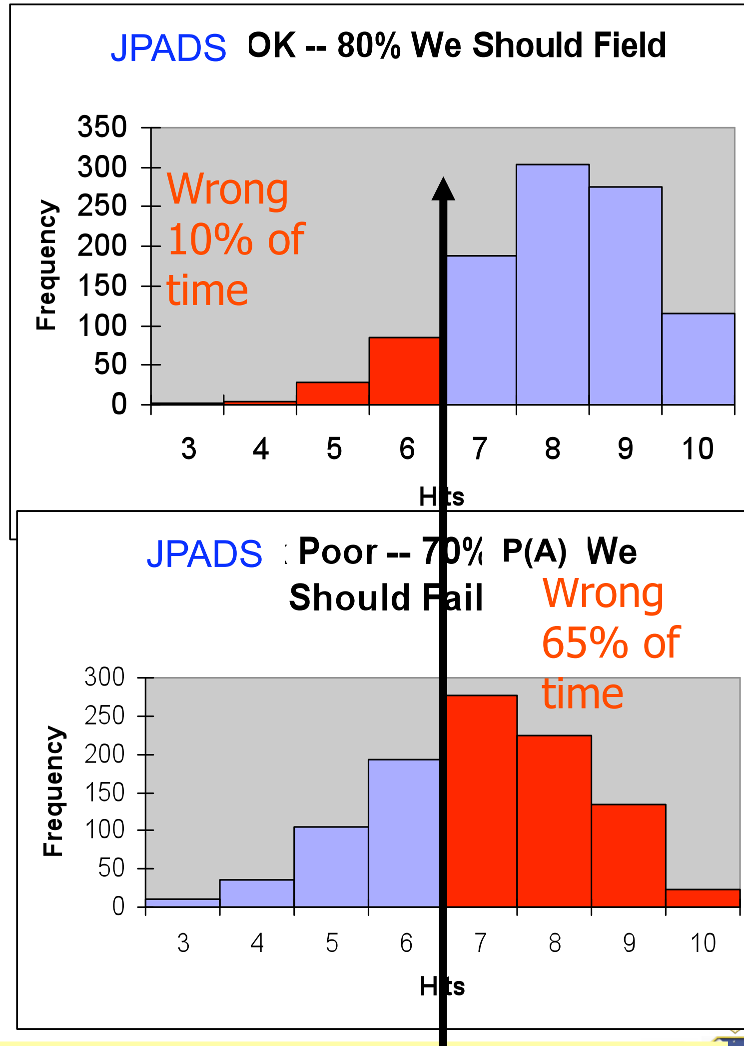
β - false negative error rate - failing to detect a difference when a difference is causally-based

Power is 1- β.

We replicate to overcome sampling error but fail to quantify the *uncertainty* in our estimates.

We seek to balance our chance of (Type I and II) errors

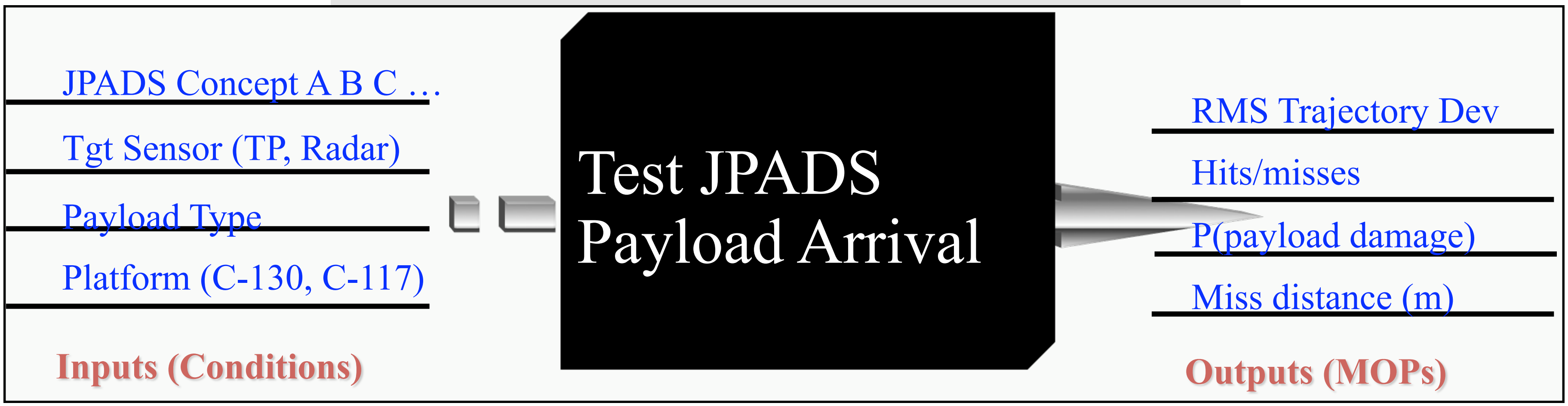
- Combining, we can trade one error for other (α for β)
- We can also increase sample size to decrease our risks in testing
- These statements not opinion –<u>mathematical fact</u> and an inescapable challenge in testing
- There are two *other* ways out … factorial designs and real-valued MOPs



Enough to Get It Right: ***Confidence*** in stating no faults; ***Power*** to detect important differences

Question 2: Which Points? How Designed Experiments Solve This

Designed Experiment *(n). Purposeful control of the inputs (factors) in such a way as to deduce their relationships (if any) with the output (responses).*



Statistician G.E.P Box said …

"All math models are false …but some are useful."

"All experiments are designed … most, poorly."

Battlespace Conditions for JPADS Case

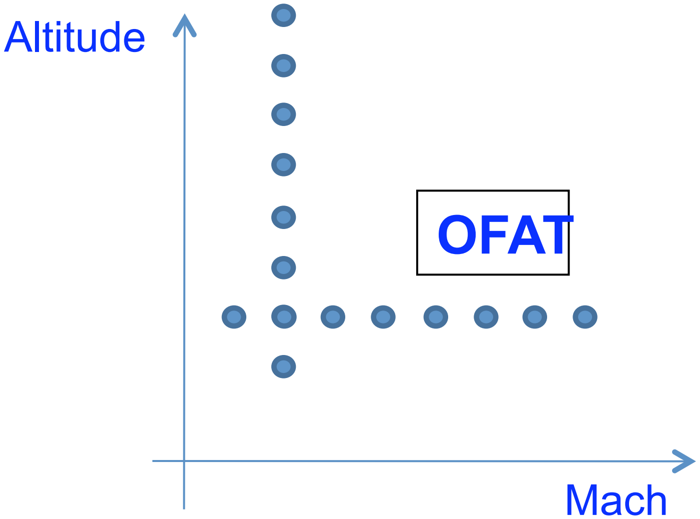
- Systems Engineering Question: Does JPADS perform at required capability level across the planned battlespace?

Type	Measure of Performance
Objective	Target acquisition range
	Target Standoff (altitude)
	launch range
	mean radial arrival distance
	probability of damage
	reliability
Subjective	Interoperability
	human factors
	tech data
	support equipment
	tactics

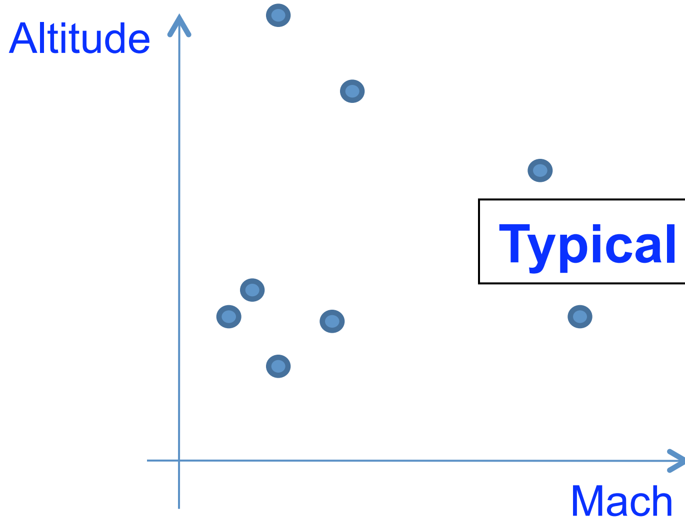
Conditions	Settings	# Levels
JPADS Variant:	A, B, C, D	4
Launch Platform:	C-130, C-17, C-5	3
Launch Opening	Ramp, Door	2
Target:	Plains, Mountain	2
Time of Day:	Dawn/Dusk, Mid-Day	3
Environment:	Forest, Desert, Snow	3
Weather:	Clear (+7nm), Haze (3-7nm), Low Ceiling/Visibility (<3000/3nm)	3
Humidity:	Low (<30%), Medium (31-79%), High (>80%)	3
Attack Azimuth:	Sun at back, Sun at beam, Sun on nose	3
Attack Altitude:	Low (<5000'), High (>5000')	2
Attack Airspeed:	Low (Mach .5), Medium (Mach .72), High (Mach .8)	3
JPADS Mode:	Autonomous, Laser Guidance	2
	Combinations	**139,968**

12 Dimensions - Obviously a large test envelope … how to search it?

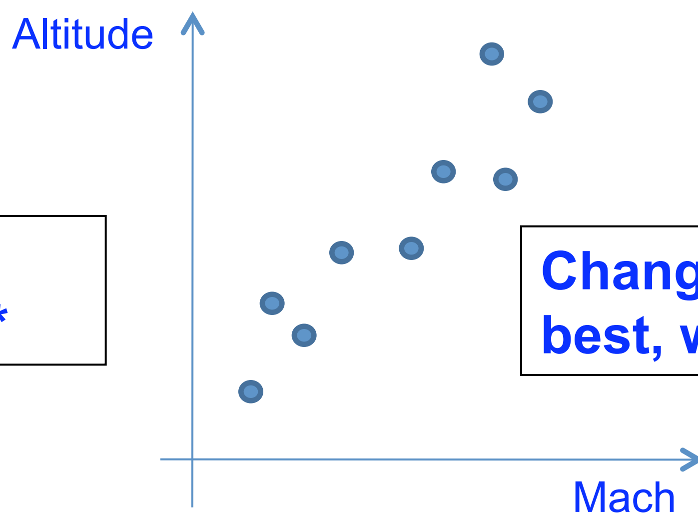
Spanning the Battlespace – Traditional Test Designs



OFAT



Typical Use Cases

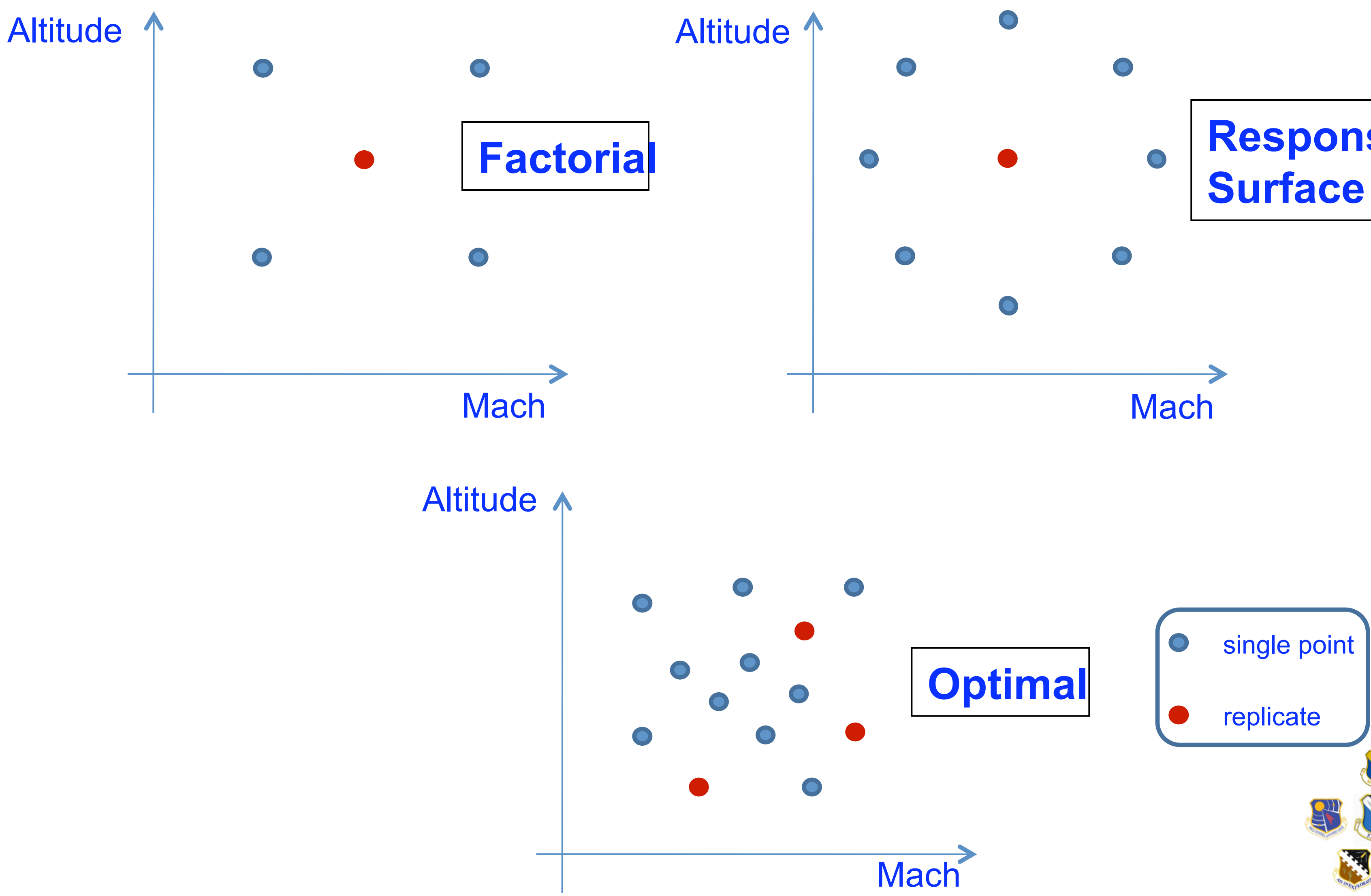


And … the always popular DWWDLT*

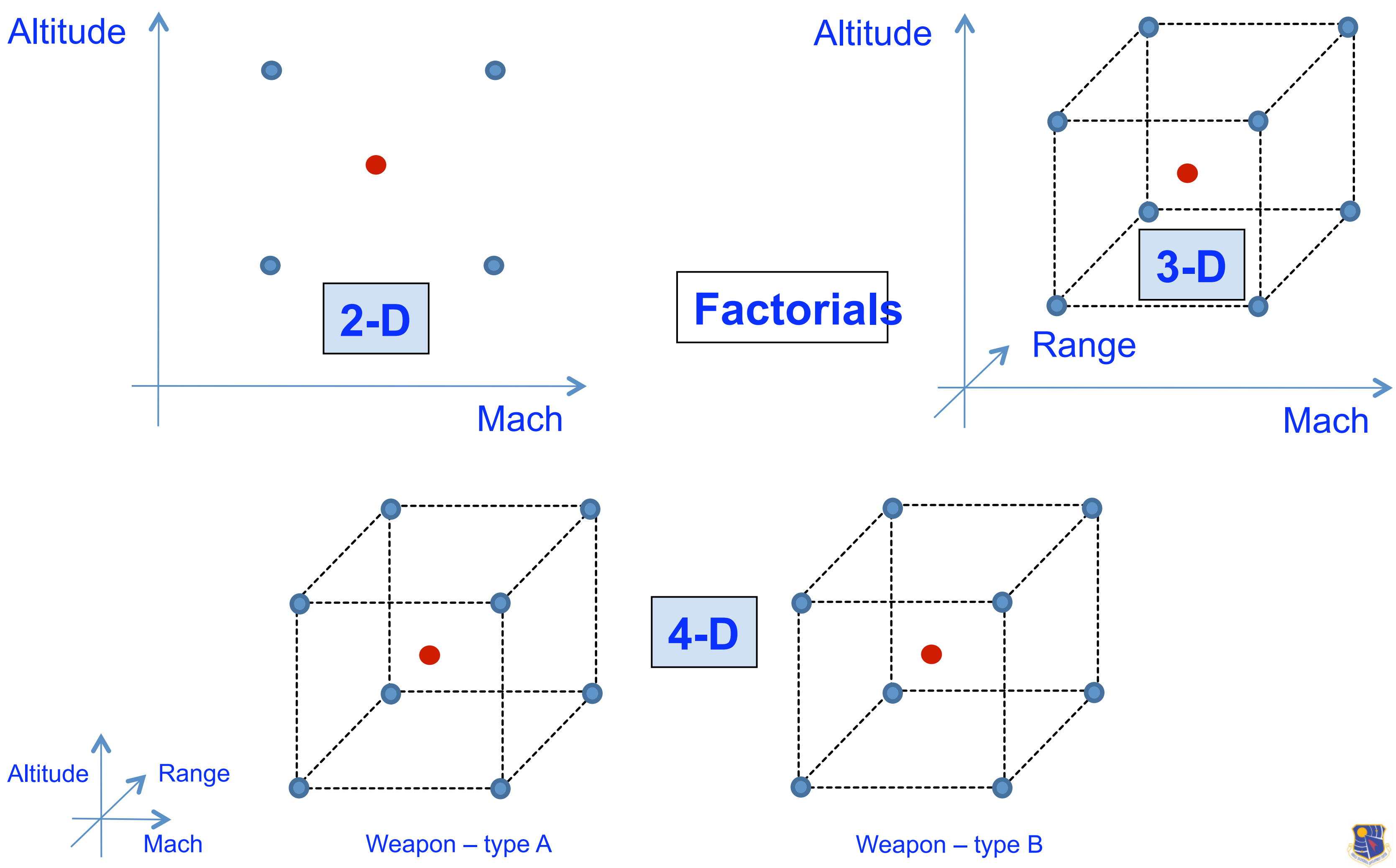
Change variables together: best, worst, nominal

*** Do What We Did Last Time**

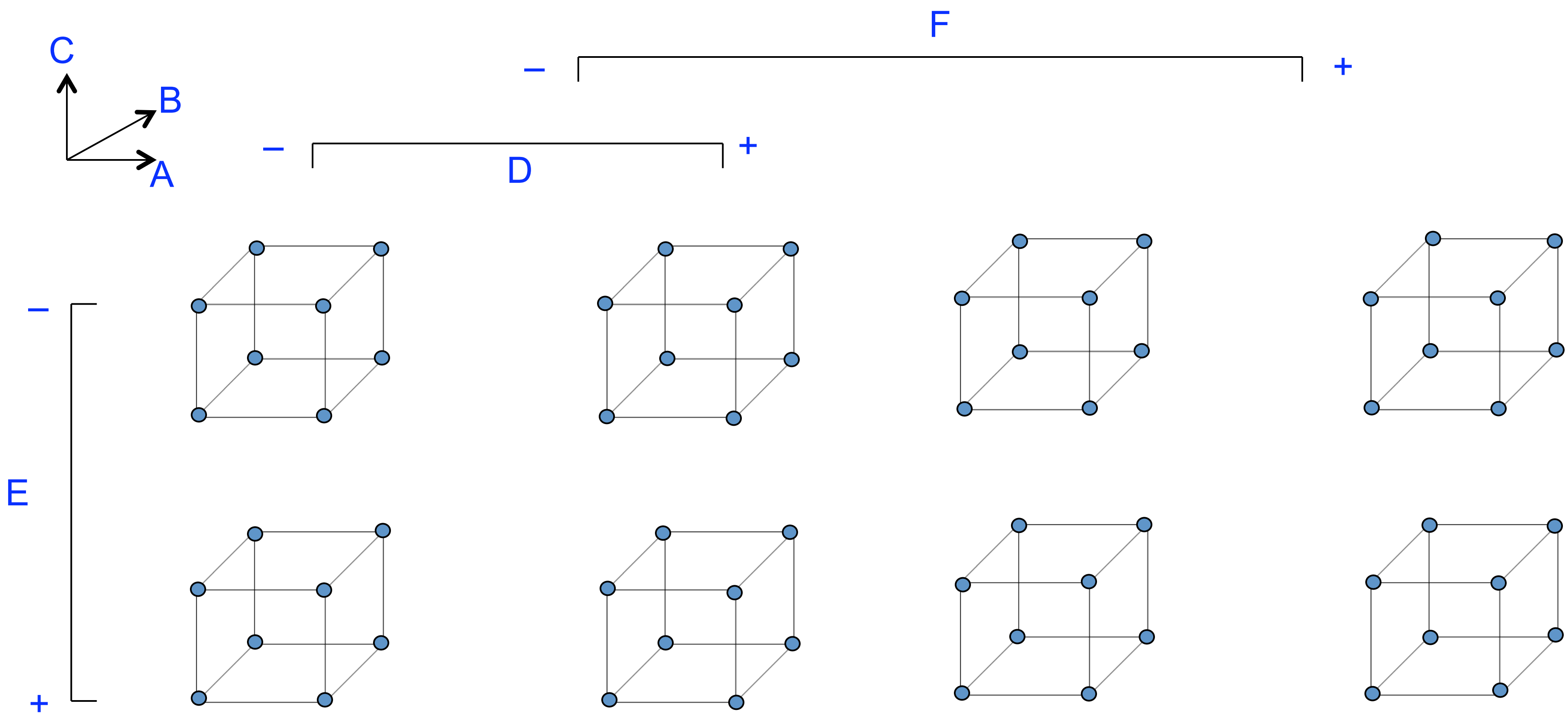
Spanning the Battlespace - DOE



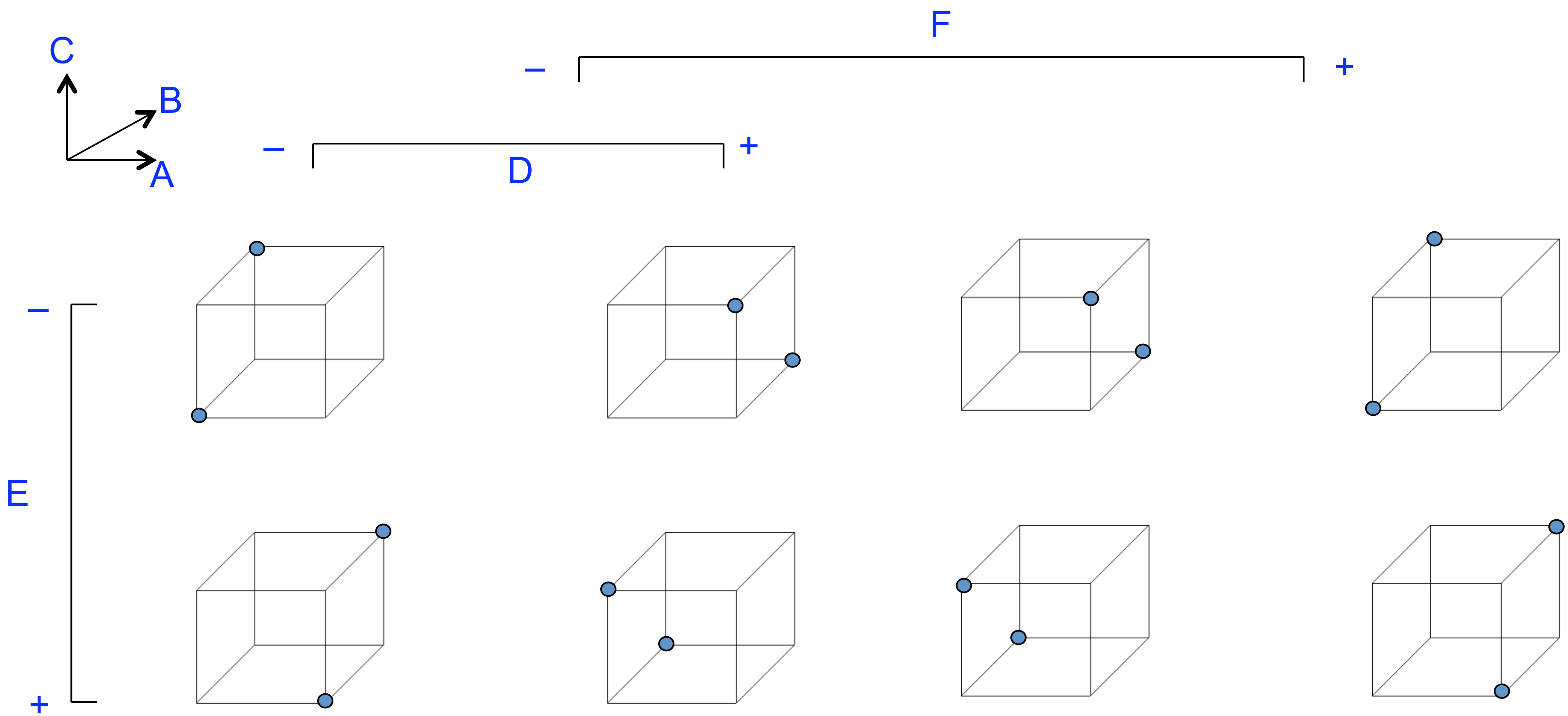
More Variables – DOE Factorials



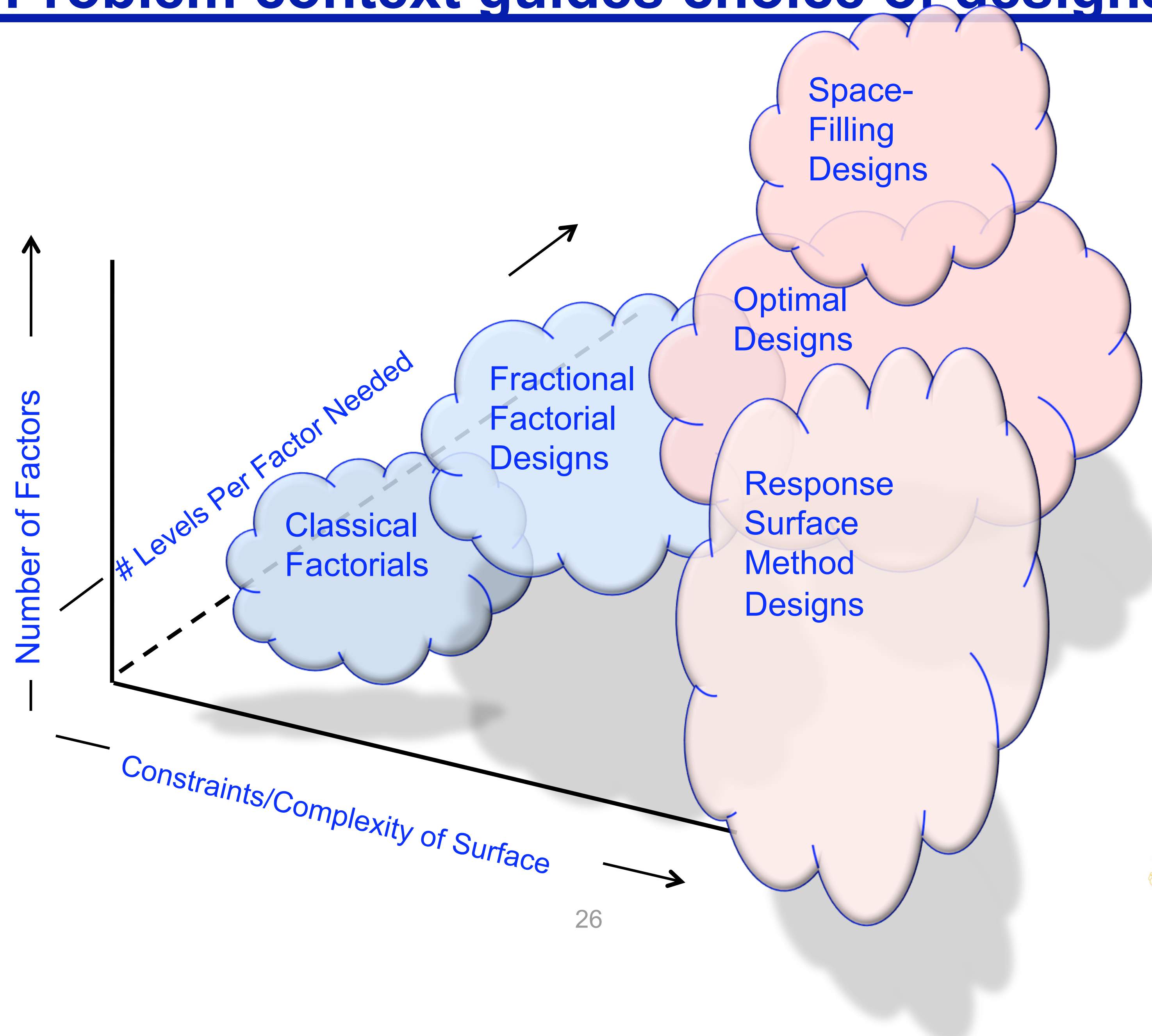
Even More Variables (here – 6)



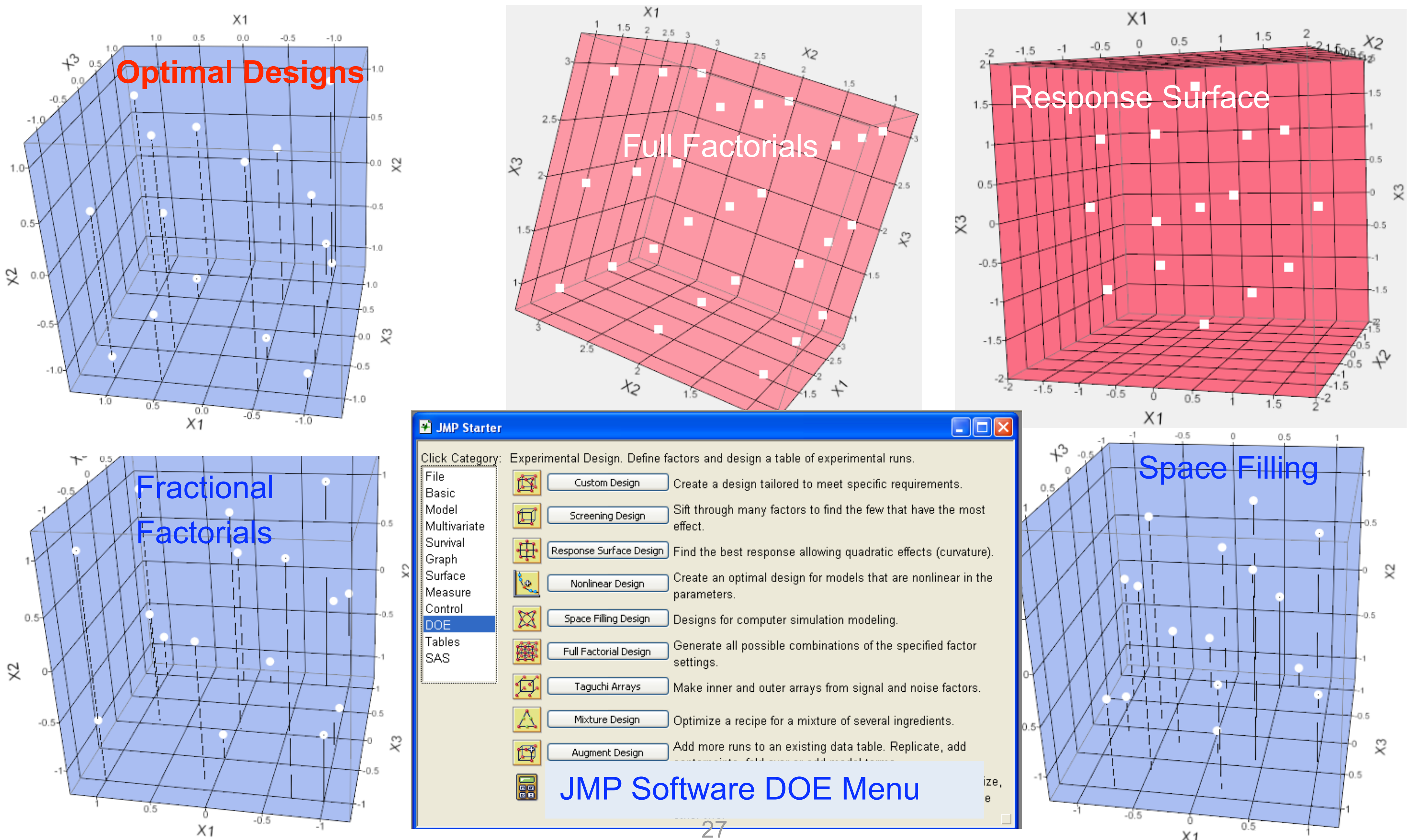
Efficiencies in Test - Fractions



Problem context guides choice of designs



We have a wide menu of design choices with DOE



Which Points to Span the Relevant Battlespace?

4 reps 1 var

JPADS A	JPADS B
4	4

2 reps 2 vars

	JPADS A	JPADS B
Ammo	2	2
Food	2	2

1 reps 3 vars

		JPADS A	JPADS B
Eglin (Low)	Ammo	1	1
	Food	1	1
Nellis (High)	Ammo	1	1
	Food	1	1

½ rep 4 vars

			JPADS A	JPADS B
Dawn (low light)	Eglin (Low)	Ammo	1	
		Food		1
	Nellis (High)	Ammo		1
		Food	1	
Midday (bright)	Eglin (Low)	Ammo		1
		Food	1	
	Nellis (High)	Ammo	1	
		Food		1

- Factorial (crossed) designs let us *learn more* from the same number of assets
- We can also use Factorials to *reduce assets* while maintaining confidence and power
- Or we can *combine* the two

All four Designs share the same ***power*** and ***confidence***

- How to support such an amazing claim?

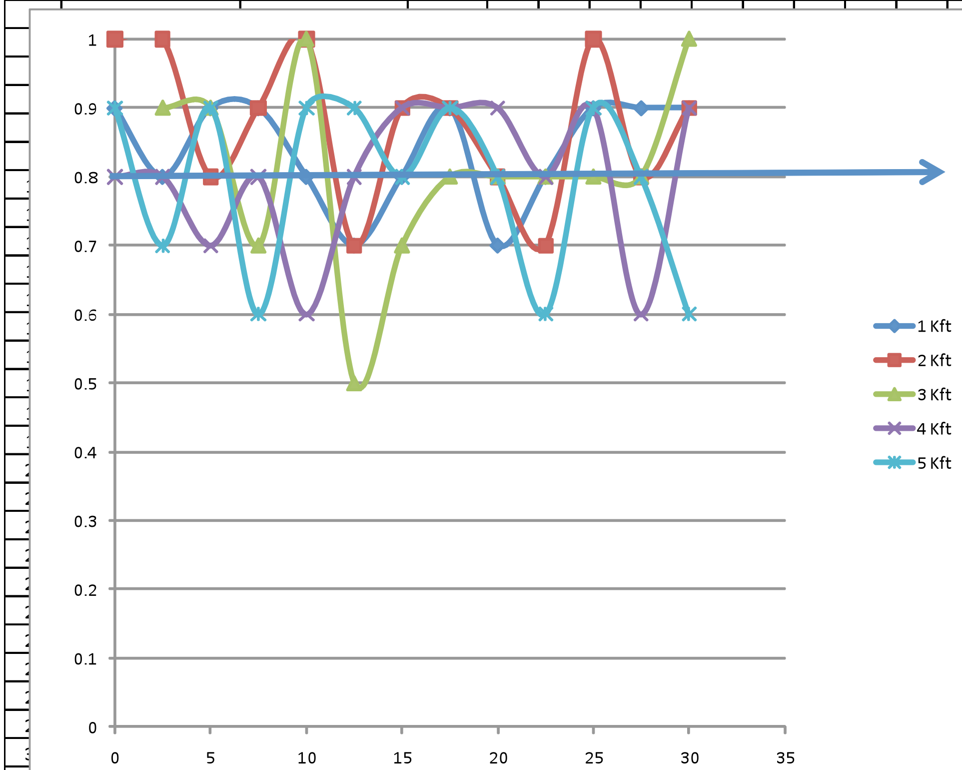
=> Switch to Excel File – JPADS Pancake.xls

Equal Power? A preposterous claim … how to justify it?

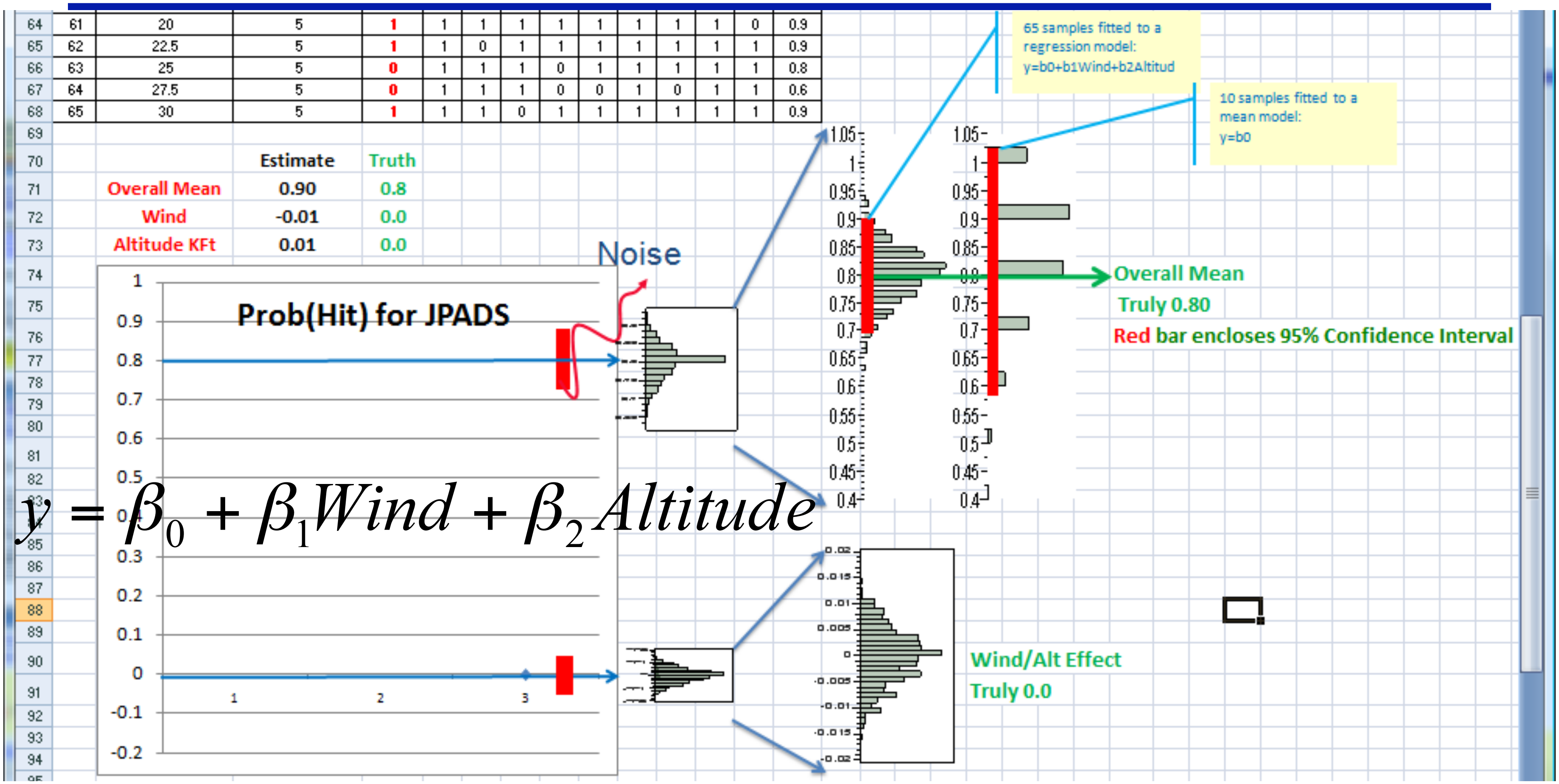
- Consider again our JPADS problem across 2 dimensions
- 13 wind speeds x 5 altitudes = 65 cases x 10 reps each = 650 trials
- Surely this will solve our problem with noise?

It will **not** … we have 65 separate 10-sample trials

			Replicates per Shot Condition										
Case	Wind	Altitude KFt	1	2	3	4	5	6	7	8	9	10	Phit
1	0	1	1	1	1	1	1	1	1	1	1	0	0.9
2	2.5	1	1	1	1	1	1	1	0	1	1	0	0.8
3	5	1	1	1	0	1	1	1	1	1	1	1	0.9
31	10	3	1	1	1	1	1	1	1	1	1	1	1
32	12.5	3	0	0	0	1	1	1	1	0	1	0	0.5
33	15	3	1	1	1	1	0	1	0	1	0	1	0.7
34	17.5	3	1	1	1	0	1	1	1	1	0	1	0.8



But, discard 9/10th of trials … strap 1/10th into a math model



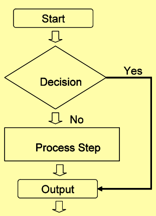
$$y = \beta_0 + \beta_1 Wind + \beta_2 Altitude$$

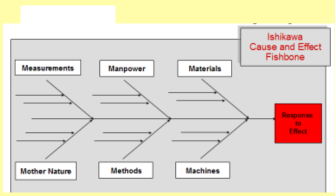
DOE math model straps all the physics together:
- *reducing* samples per condition by 90% while
- *increasing* our prediction accuracy 50%

Note: this speaks to the method of analysis (Challenge #4.)

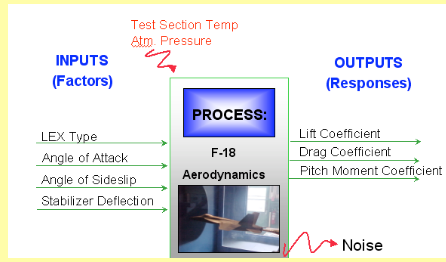
Test as Science vs. Art: Experimental Design
Test Process is Well-Defined

Planning: Factors Desirable and Nuisance

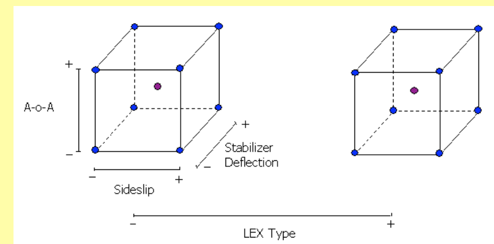




Desired Factors and Responses

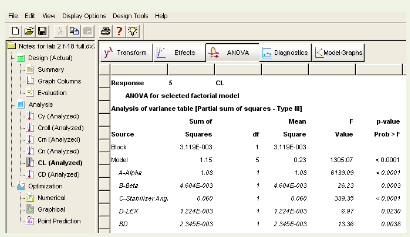


Design Points



Test Matrix

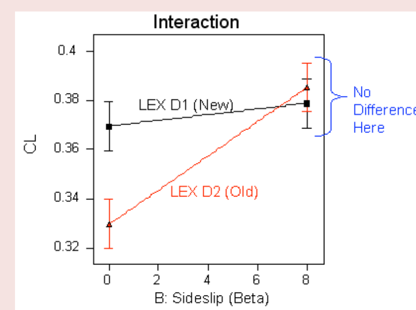
Randomize & Block -> Results and Analysis



Model Build

C_L =
+0.38
+0.26 x A-o-A
+0.017 x Sideslip
+0.061 x Stabilizer Deflection
-.00875 x LEX Type
+0.012 x Sideslip x LEX Type

Discovery, Understanding Prediction, Re-design



It applies to our tests: DOE in 50+ operations over 20 years

- IR Sensor Predictions
- Ballistics 6 DOF Initial Conditions
- Wind Tunnel fuze characteristics
- Camouflaged Target JT&E ($30M)
- AC-130 40/105mm gunfire CEP evals
- AMRAAM HWIL test facility validation
- 60+ ECM development + RWR tests
- GWEF Maverick sensor upgrades
- 30mm Ammo over-age LAT testing
- Contact lens plastic injection molding
- 30mm gun DU/HEI accuracy (A-10C)
- GWEF ManPad Hit-point prediction
- AIM-9X Simulation Validation
- Link 16 and VHF/UHF/HF Comm tests
- TF radar flight control system gain opt
- New FCS software to cut C-17 PIO
- AIM-9X+JHMCS Tactics Development
- MAU 169/209 LGB fly-off and eval
- Characterizing Seek Eagle Ejector Racks
- SFW altimeter false alarm trouble-shoot
- TMD safety lanyard flight envelope
- Penetrator & reactive frag design
- F-15C/F-15E Suite 4 + Suite 5 OFPs
- PLAID Performance Characterization
- JDAM, LGB weapons accuracy testing
- Best Autonomous seeker algorithm
- SAM Validation versus Flight Test
- ECM development ground mounts (10's)
- AGM-130 Improved Data Link HF Test
- TPS A-G WiFi characterization
- MC/EC-130 flare decoy characterization
- SAM simulation validation vs. live-fly
- Targeting Pod TLE estimates
- Chem CCA process characterization
- Medical Oxy Concentration T&E
- Multi-MDS Link 16 and Rover video test

Adopt a Policy of Well-Designed Tests

46TW DOE policy.pdf - Adobe Reader

File Edit View Document Tools Window Help

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS 46th TEST WING (AFMC)
101 WEST D AVE SUITE 226
EGLIN AIR FORCE BASE FLORIDA 32542-5000

7 Jul 09

POLICY LETTER FOR ALL 46 TW PERSONNEL

FROM: 46 TW/CC

SUBJECT: Design of Experiments (DOE) is 46 TW Primary Test Strategy

1. Under AFSO21, our Chief challenged the Air Force Test and Evaluation (T&E) Enterprise to make T&E more effective and efficient. For the past several years, we've monitored our sister test organizations as they applied the principles of DOE as their primary test strategy. During 2007-09, we engaged in a robust "DOE proof of concept" phase spanning more than 40 projects throughout our test portfolio. Trials are concluded; DOE works! We can improve 46 TW tests by adopting the principles of the science of test in every program where it makes sense.

2. Therefore, each Test Squadron will use DOE in all their testing when they have control of test design and when the number of test events consists of more than a mere demonstration. Exceptions will be approved by the appropriate Group Commander.

3. All 46 TW-designed test plans will 1) mathematically cite the statistical risks implied by their proposed test program, and 2) achieve high confidence and power over a broad test volume. Suitable modifications to this policy will be made for software tests. Policy compliance at the Group (or squadron level where appropriate) will be tracked quarterly at Wing Staff meetings using the metrics in attachment 1. The office of the Wing Operations Analyst is the primary technical point of contact for this policy.

Checklist: Fruits of Well-Designed Tests

- ❑Specify Goal/Objective
- ❑List Quantitative Responses
- ❑List factors/levels & how to control in test
- ❑Strategy to place Points
- ❑Compute Confidence/ Power

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST AND EVALUATION

OCT 1 9 2010

MEMORANDUM FOR COMMANDER, ARMY TEST AND EVALUATION COMMAND
COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE
COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND
DEPUTY UNDER SECRETARY OF THE ARMY, TEST & EVALUATION COMMAND
DEPUTY, DEPARTMENT OF THE NAVY TEST & EVALUATION EXECUTIVE
DIRECTOR, TEST & EVALUATION, HEADQUARTERS, U.S. AIR FORCE
TEST AND EVALUATION EXECUTIVE, DEFENSE INFORMATION SYSTEMS AGENCY
DOT&E STAFF

SUBJECT: Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation

This memorandum provides further guidance on my initiative to increase the use of scientific and statistical methods in developing rigorous, defensible test plans and in evaluating their results. As I review Test and Evaluation Master Plans (TEMPs) and Test Plans, I am looking for specific information. In general, I am looking for substance vice a 'cookbook' or template approach - each program is unique and will require thoughtful tradeoffs in how this guidance is applied.

A "designed" experiment is a test or test program, planned specifically to determine the effect of a factor or several factors (also called independent variables) on one or more measured responses (also called dependent variables). The purpose is to ensure that the right type of data and enough of it are available to answer the questions of interest. Those questions, and the associated factors and levels, should be determined by subject matter experts -- including both operators and engineers -- at the outset of test planning.

Design of Experiments is a structured process to identify the metrics, factors, and levels that most directly affect operational effectiveness and suitability and that should be reflected in detailed test plans. DOT&E is working with other members of the test and evaluation community to develop a two-year roadmap for implementing this scientific and rigorous approach to testing. I am looking for as much substance as possible as early as possible, but each TEMP revision can be tailored as more information becomes available. That content can either be explicitly made part of TEMPs and Test Plans, or referenced in those documents and provided separately to DOT&E for review.

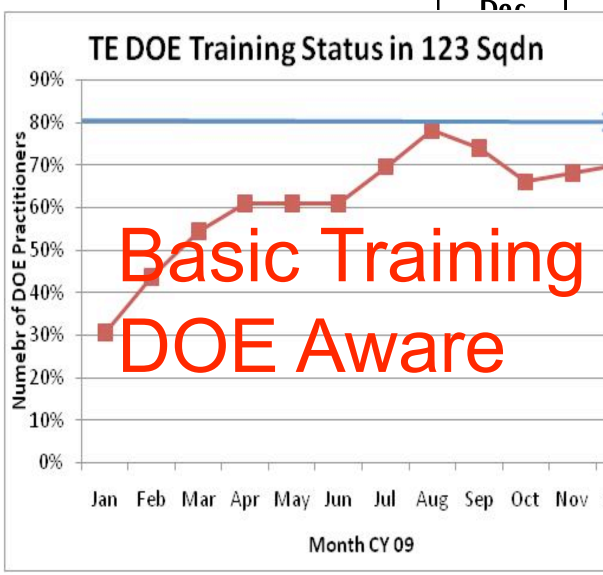
J. M. Gilmore

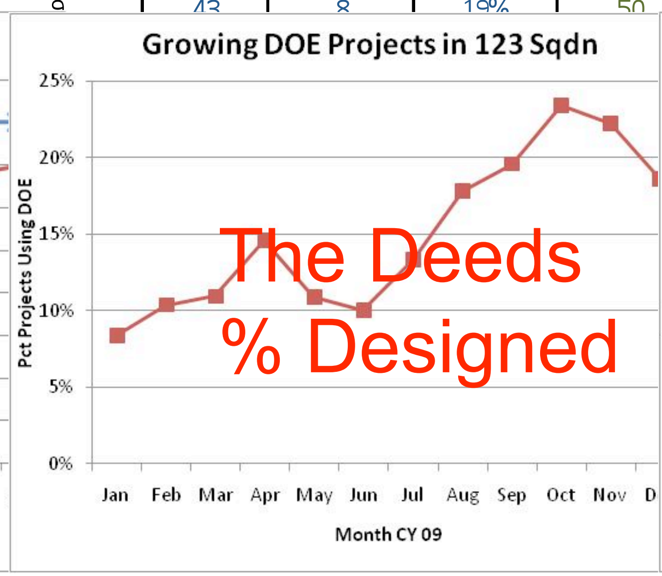
J. Michael Gilmore
Director

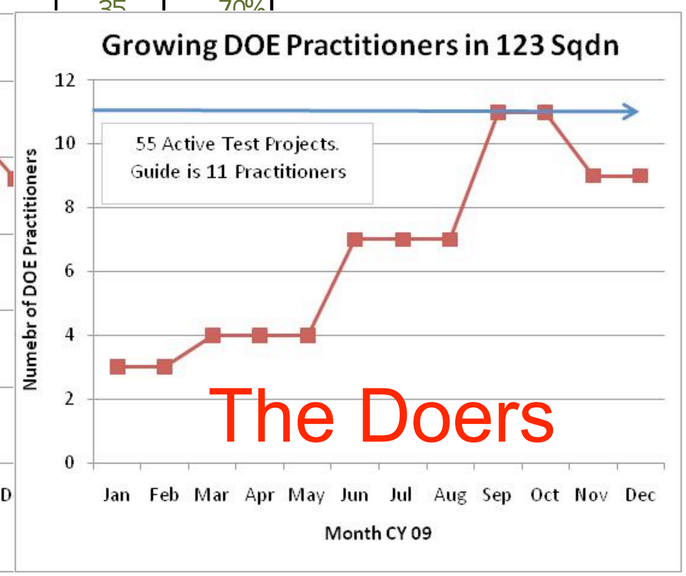
cc:
DDT&E

What you measure gets done …
Sample Unit Quarterly Metrics

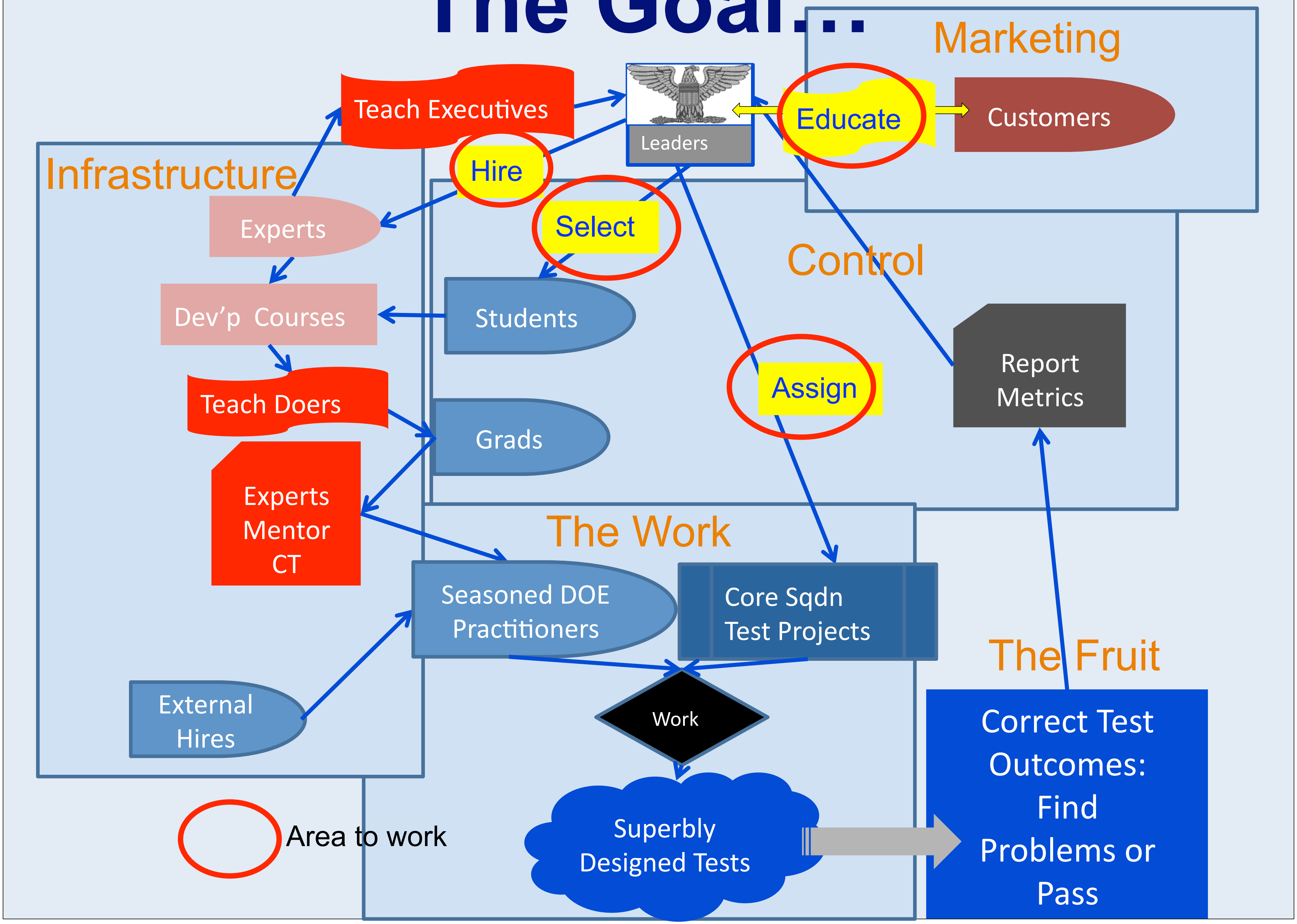
DOE Metrics Table							
Month	Practitioners	Active Projects	DOE Projects	% DOE Projects	Assigned PE/TE	DOE-Trained	% DOE-Trained
Jan	3	60	5	8%	46	14	30%
Feb	3	58	6	10%	46	20	43%
Mar	4	55	6	11%	46	25	54%
Apr	4	48	7	15%	46	28	61%
May	4	46	5	11%	46	28	61%
Jun	7	40	4	10%	46	28	61%
Jul	7	45	6	13%	46	32	70%
Aug	7	45	8	18%	46	36	78%
Sep	11	46	9	20%	50	37	74%
Oct	11	47	11	23%	50	33	66%
Nov	9	45	10	22%	50	34	68%
Dec	9	43	8	19%	50	35	70%







The Goal…



Marketing
Teach Executives
Leaders
Educate
Customers
Infrastructure
Hire
Experts
Select
Control
Dev'p Courses
Students
Report
Metrics
Teach Doers
Assign
Grads
Experts
Mentor
CT
The Work
Seasoned DOE
Practitioners
Core Sqdn
Test Projects
The Fruit
External
Hires
Work
Correct Test
Outcomes:
Find
Problems or
Pass
Area to work
Superbly
Designed Tests

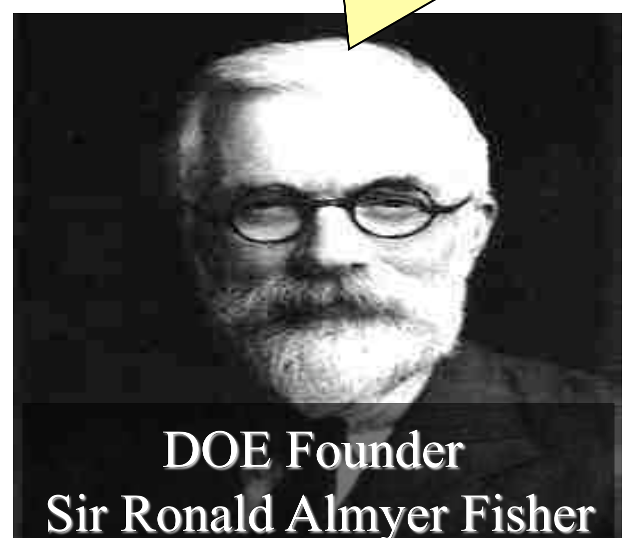
In Memorium R.A. Fisher

- Principles of DOE
  - <Orthogonality>
  - Randomization
  - Replication
  - Local Control of Error

"To call in the statistician after the experiment is . . . asking him to perform a postmortem examination: he may be able to say what the experiment died of."

Address to Indian Statistical Congress, 1938.

"No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question at a time. The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed." R. A. Fisher



DOE Founder
Sir Ronald Almyer Fisher

So, What's the Good News?

We Have ***Great*** Answers to ***Key*** Questions.
- It's the way we build better tests
- N, points, order, conclusions?
- Uniquely answers deep and broad challenges
- Quantify the test risks DOD incurs
- Less-experienced testers can reliably succeed

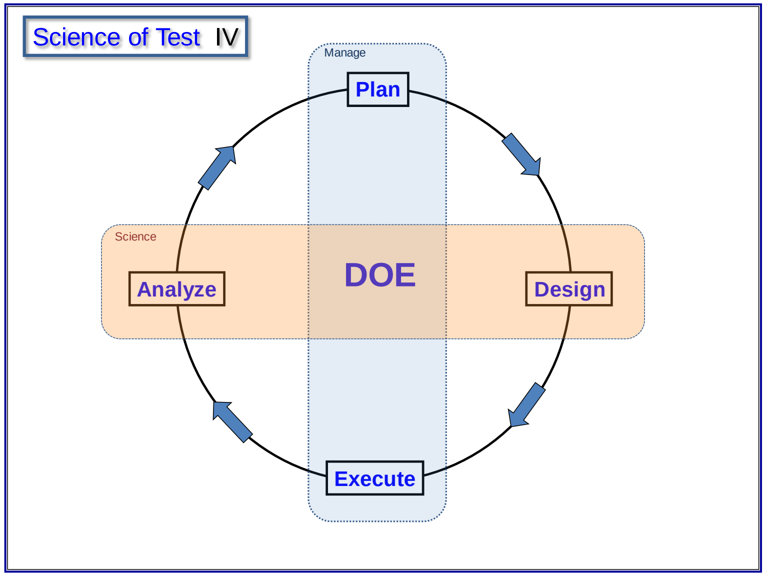
- Small town Ga quarterback…
- A final challenge … Lead us!

George Harrison, MGen USAF (ret)

What's *Your* Method of Test?

DOE: The *Science* of Test

Questions?



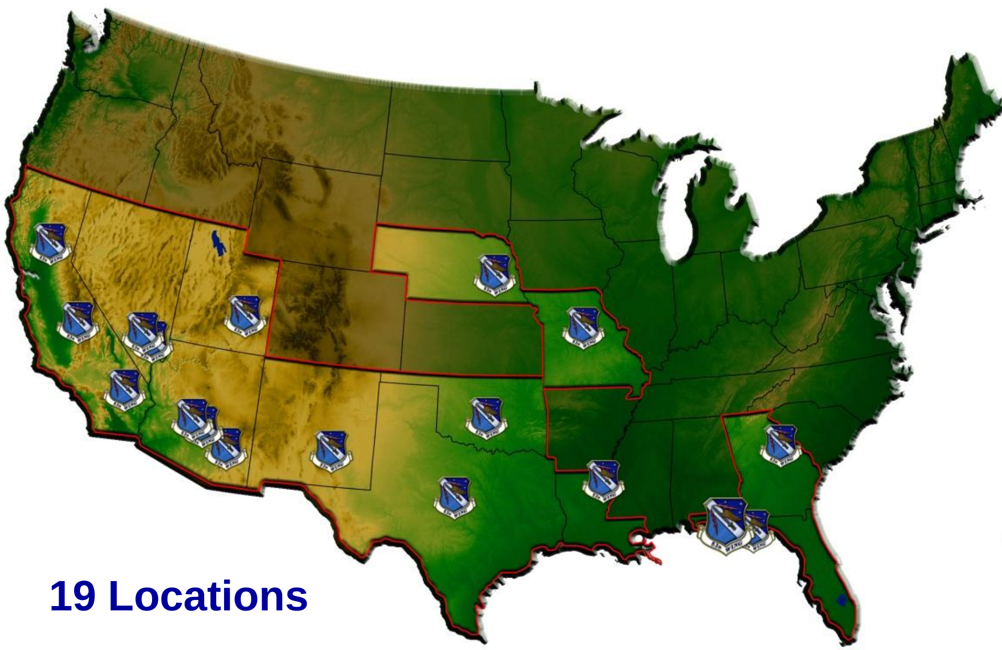
Embedding DOE in Military Testing
One Organization's Roadmap

presented to:
2011 NASA Statistical Engineering Symposium
May 2011

Jim Simpson, 53d Wing – Greg Hutto 46 Test Wing – Alex Sewell 53d Wing

53d Wing

Mission: Develop, test and evaluate advanced weapons, systems and tactics to perfect the lethality, survivability, and sustainability of our nation’s combat forces



19 Locations

2000 Professionals consisting of…

- **550 Officers**
- **650 Enlisted**
- **450 Civilians**
- **350 Contractor**

53d Wing
Analysts and Test Engineers

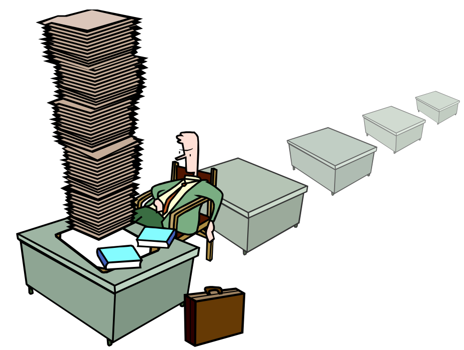
53d Wing

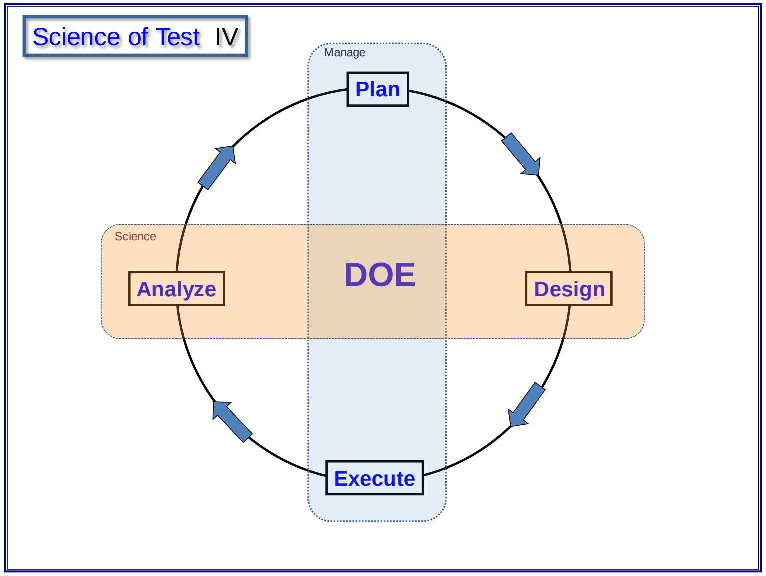
	Electronic Warfare	Weapons Evaluation	Test & Evaluation	Test Management
Analysts 71	7	9	7	48
Test Engineers 124	50	16	20	38

What is *Your* Dream?

"Be Careful What you Ask for . . ." **Kevin Burns, Ops Test, Tech Advisor**

DOE is Policy
In DoD





Changing a Culture

Contrast Traditional Methods …

Cases

Case	Configuration	Outcome
1		Good
2		Good
3		OK
4		Good
5		Good

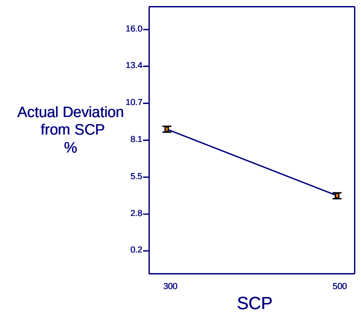
Good to go!

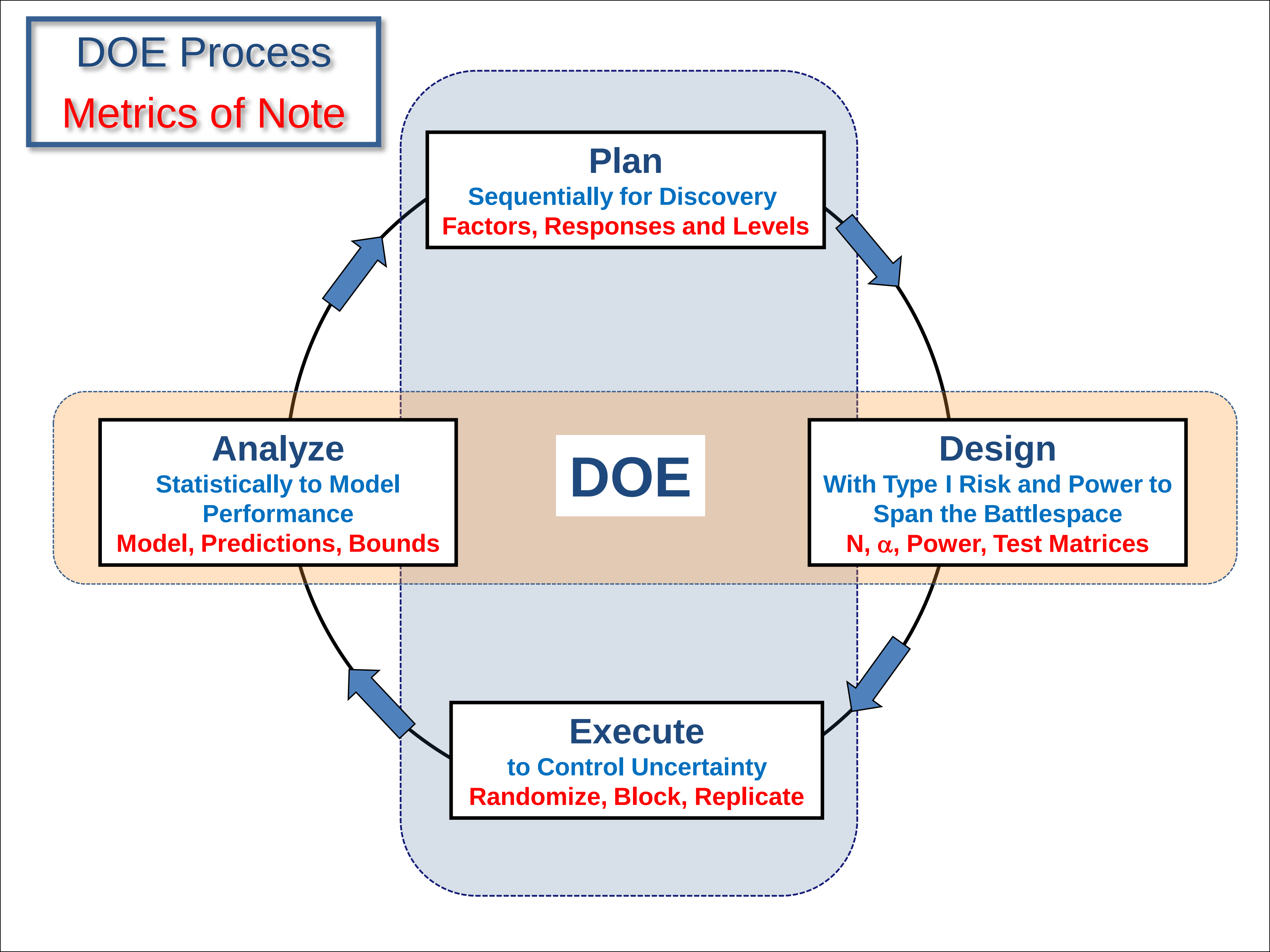
OR

One Factor-at-a-Time

Case	A	B	C
1	**1**	0	0
2	**2**	0	0
3	**3**	0	0
4	**4**	0	0
5	0	**1**	0
6	0	**2**	0
7	0	**3**	0
8	0	**4**	0
9	0	0	**1**
10	0	0	**2**
11	0	0	**3**
…	…	…	…

Effect Graph



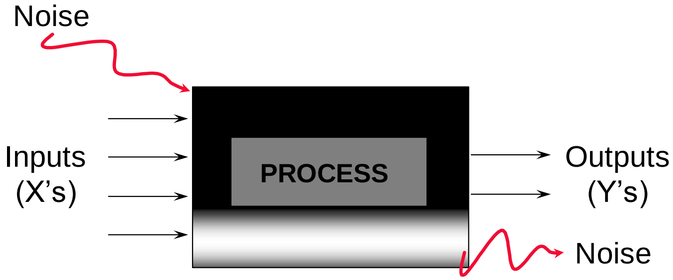

DOE Process
Metrics of Note
Plan
Sequentially for Discovery
Factors, Responses and Levels
Design
With Type I Risk and Power to Span the Battlespace
N, α, Power, Test Matrices
Execute
to Control Uncertainty
Randomize, Block, Replicate
Analyze
Statistically to Model Performance
Model, Predictions, Bounds
DOE

Questions in Testing

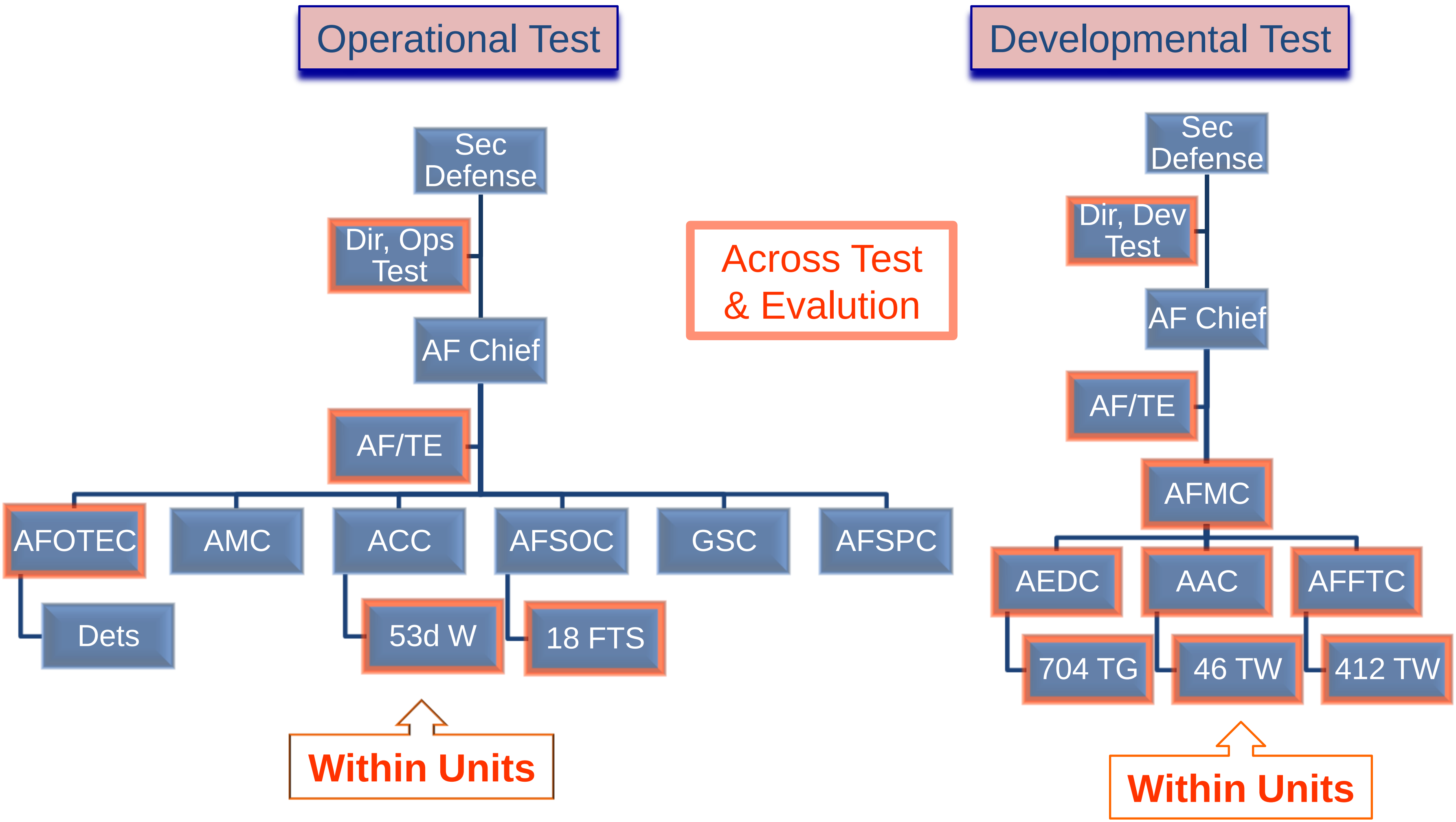
Four Challenges faced by any test

1. *How Many?* A: Sufficient samples to control our twin errors – false positives & negatives
2. *Which Points and What's Good?* A: Span the battle-space with orthogonal run matrices using continuous measures tied to the test objectives
3. *How Execute?* A: Randomize and block runs to exclude effects of the lurking, uncontrollable nuisance variation
4. *What Conclusions?* A: Build math-models of input/output relations, quantifying noise, controlling error

Design of Experiments effectively addresses all these challenges!

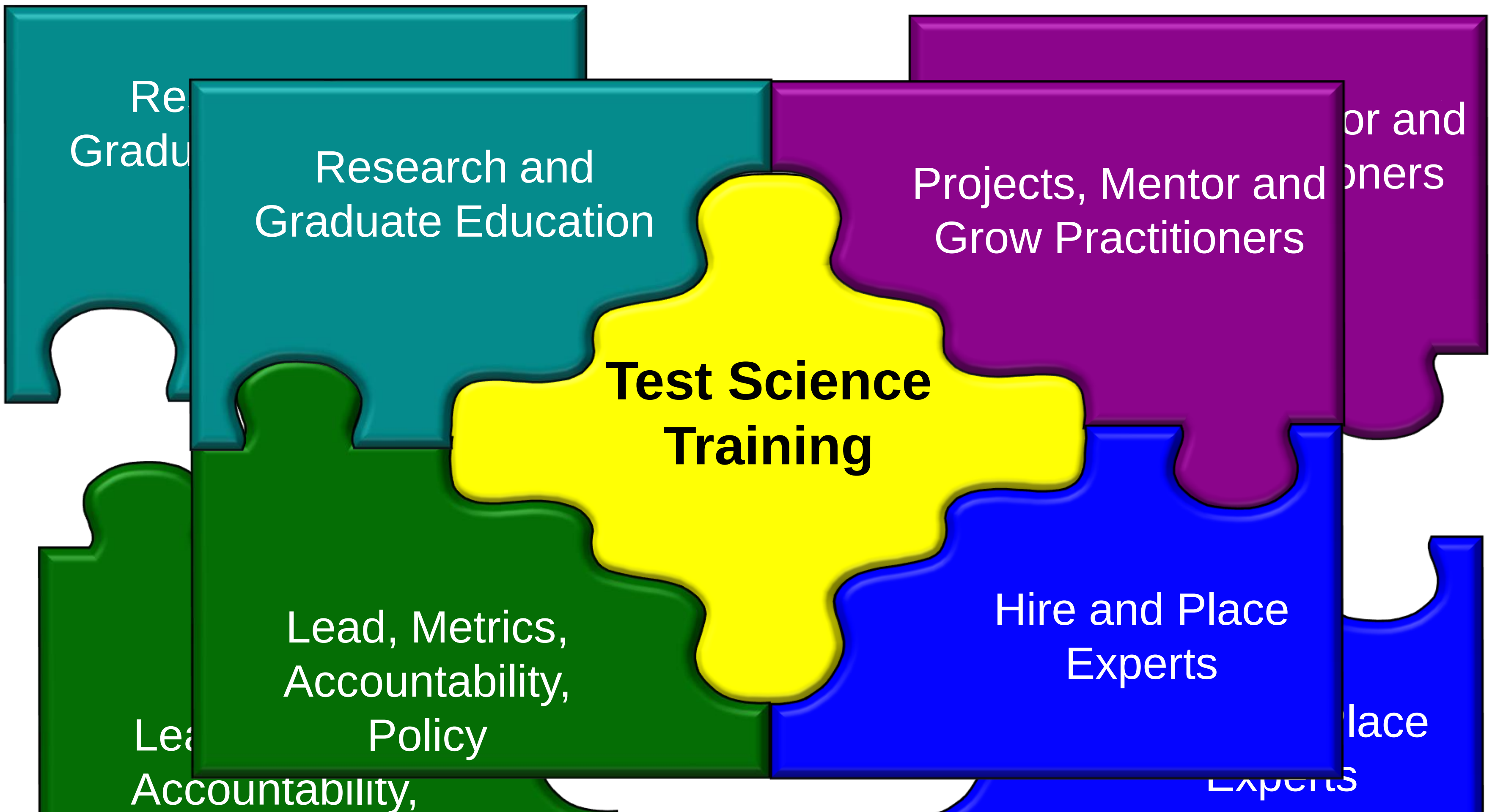


Culture Change Across Units



Organization Change Pieces

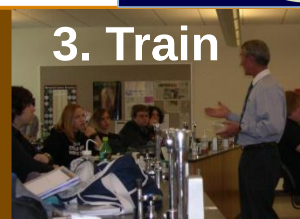
Move into Place Simultaneously



- Requires integrated effort with emphasis on each piece to pull it all together to affect the way we test

Science of Test
Steps to Implementation within Unit

5. Standards



3. Train

4. Mentor

2. Short-Term Wins

I. Leadership --Why DOE?

II. Technical Continuity

1. Foundations

III. Process Improvement

IV. Change Org Structures

Leading the Science of Test

- Stay tuned for the next talk …

I. Leadership --Why DOE?

II. Technical Continuity

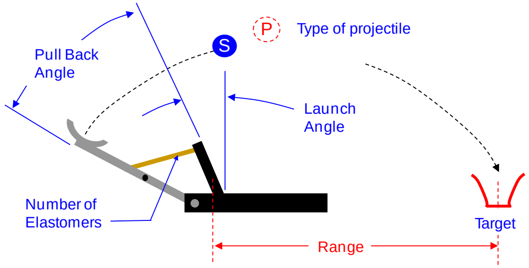
1. Foundations

III. Process Improvement

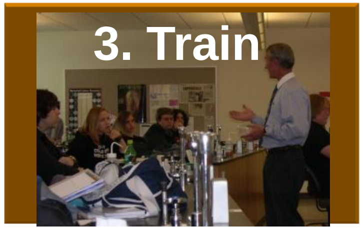
IV. Change Org Structures

Training our Total Test Team

- Leadership, Support and Operator Series
  - DOE Executive Interview (1-2 hour)
  - DOE for Leaders, Aircrew (half day)
  - Intro to Design of Experiments (2 days)
  - DOE Foundations (1 week)

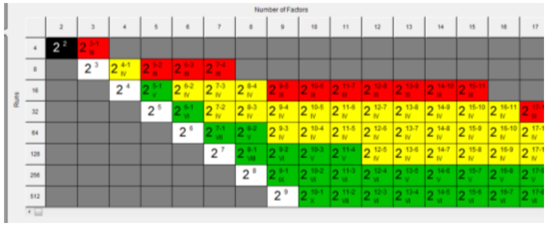


- Analyst and Test Engineer Practitioner Series
  - Each 1-week course uses Discussion-Seatwork-Projects
  - DOE 0 – DOE Foundations for Science of Test
  - DOE I – Design and Analysis of Factorial and Fractionated Designs
  - DOE II – Response Surface Methods, Optimal Designs, Split Plots, Analysis of Ugly Data

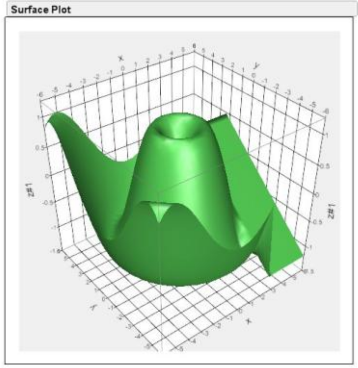


Software for Practitioners

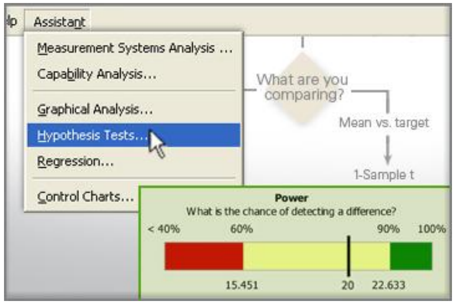
- *Design Expert* – software solely for design of experiments
  - Keeps the analyst focused on DOE procedure
  - Warns when going wayward
  - Used in DOE 0, I, II and in-part III



- *JMP* – general purpose statistical software
  - Industry leader, affordable, requires learning curve
  - Best for our advanced users and needs
  - For DOE III and difficult problems



- *Minitab* – general purpose
  - Interface similar to Excel, user friendly
  - DOE emphasis
  - Split-plot capable



Growing & Mentoring Practitioners

Practitioner* -- (prak-tish-un-ur) n. 1. One who practices an occupation, method or technique.

28 TES/EAA Initial OA Qual Training (ch8 14 Aug 09)

Trainee:	Date complete:				
Training Items	**Method**	**Training Information**	**Date Complete**	**Trainee Initials**	**Trainer Initials**
Wing Directed (mandatory) Training - (complete prior to initial qualification)					
Test Team Training (TTT)	FC	53TMG/TR Webpage			
Design of Experiment (DOE) 0	FC	53TMG/TR Webpage			
DOE I	FC	53TMG/TR Webpage			
DOE II	FC	53TMG/TR Webpage			
53WG 99-103 Review (complete prior to initial qualification)					
Read 53 WGI 99-103 *Capability Based T&E*	SS	53WG Handbook			
Test Team Handbook Review (complete prior to initial qualification)					
Review 53WG Test Team Handbook	SS	53WG Webpage			
Test process checklists	SS	53WG Handbook			
Test template review	SS	53WG Handbook			
Test regulation review	SS	53WG Handbook			
Test Management Review (complete prior to initial qualification)					
TMS Use / Procedures	OJT	PM			
Attend MRR	SS	TBD			
Attend CoT/CRR/FRR	SS	Thursdays (1300)			
Review Test Priority List	SS	TMS			

28 TES/EAA Experienced OA Qual Training (Ch8 14 Aug 09)

Trainee:	Date complete:				
Training Items	**Method**	**Training Information**	**Date Complete**	**Trainee Initials**	**Trainer Initials**
Wing Training - Experienced OA Qual (complete prior to experienced OA qualification)					
Project Management Training (PMT)	FC	TMG Webpage			
Operations Suitability Training (OST)	FC	TMG Webpage			
DOE III	FC	TMG Webpage			
Supplemental Certification Training (complete courses prior to experienced OA qualification)					
ACQ 101 - Fundamentals of Systems Acquisition Management	OL	DAU website			
SYS 101 - Systems Planning, Research, Development & Engineering	OL	DAU website			
Level I T&E Certification					
TST 102 - Fundamentals of Test & Evaluation	OL	DAU website			
CLE 023 - Modeling and Simulation for Test and Evaluation	OL	DAU website			
Level I Program Management Certification					
CLB 007 - Cost Analysis	OL	DAU website			
CLB 016 - Introduction to Earned Value					

- **Various practitioner levels – requires experience**
  - **OA – Initial Qual, Experienced, Instructor**
  - **TE – Initial Qual, Experienced**
- **Include re-qualification**

Long Term Solution
Leadership: Making Changes Endure

DOT&E, DDT&E and Service TE Policy Providing Leadership

MEMORANDUM OF AGREEMENT

SUBJECT: Using Design of Experiments for Operational Test and Evaluation

Regarding the subject, we endorse the enclosed findings of the Operational Test Agency Technical Directors and the Science Advisor for Operational Test and Evaluation.

Dr. Charles E. McQueary
Director, Operational Test & Evaluation

Stephen T. Sargeant, Major General, USAF
Commander, AFOTEC

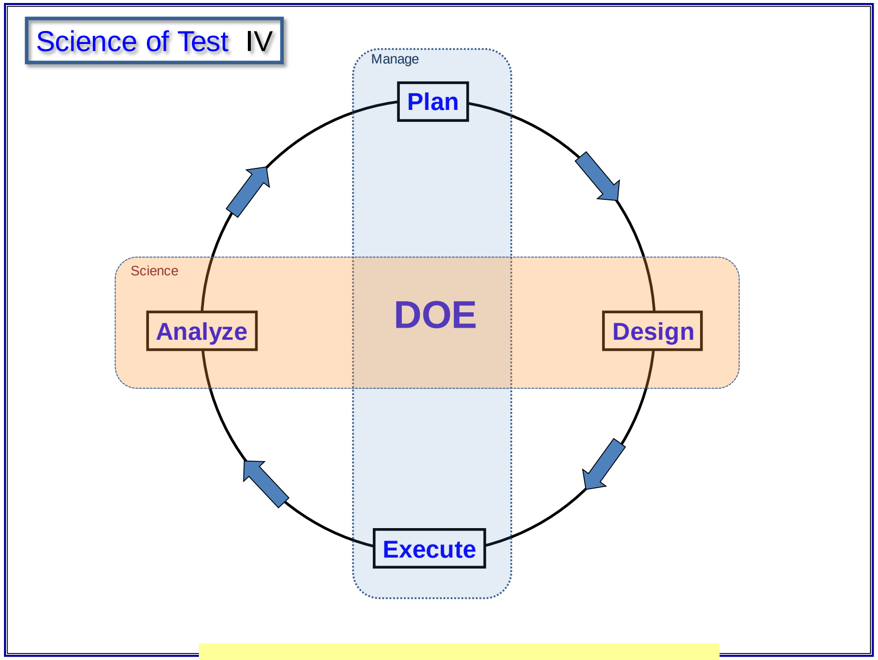
Roger A. Nadeau, Major General, USA
Commander, ATEC

David A. Dunaway, Rear Admiral, USN
Commander, OPTEVFOR

David L. Reeves, Colonel, USMC
Director, MCOTEA

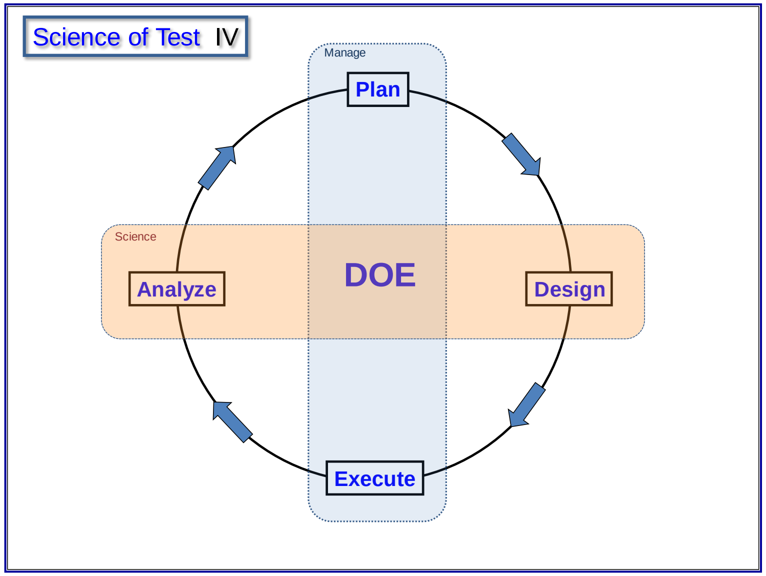
Ronald C. Stephens, Colonel, USA
Commander, JITC

Policy/Guidance



Accountability/Metrics

5. Standards



Defining What We Do

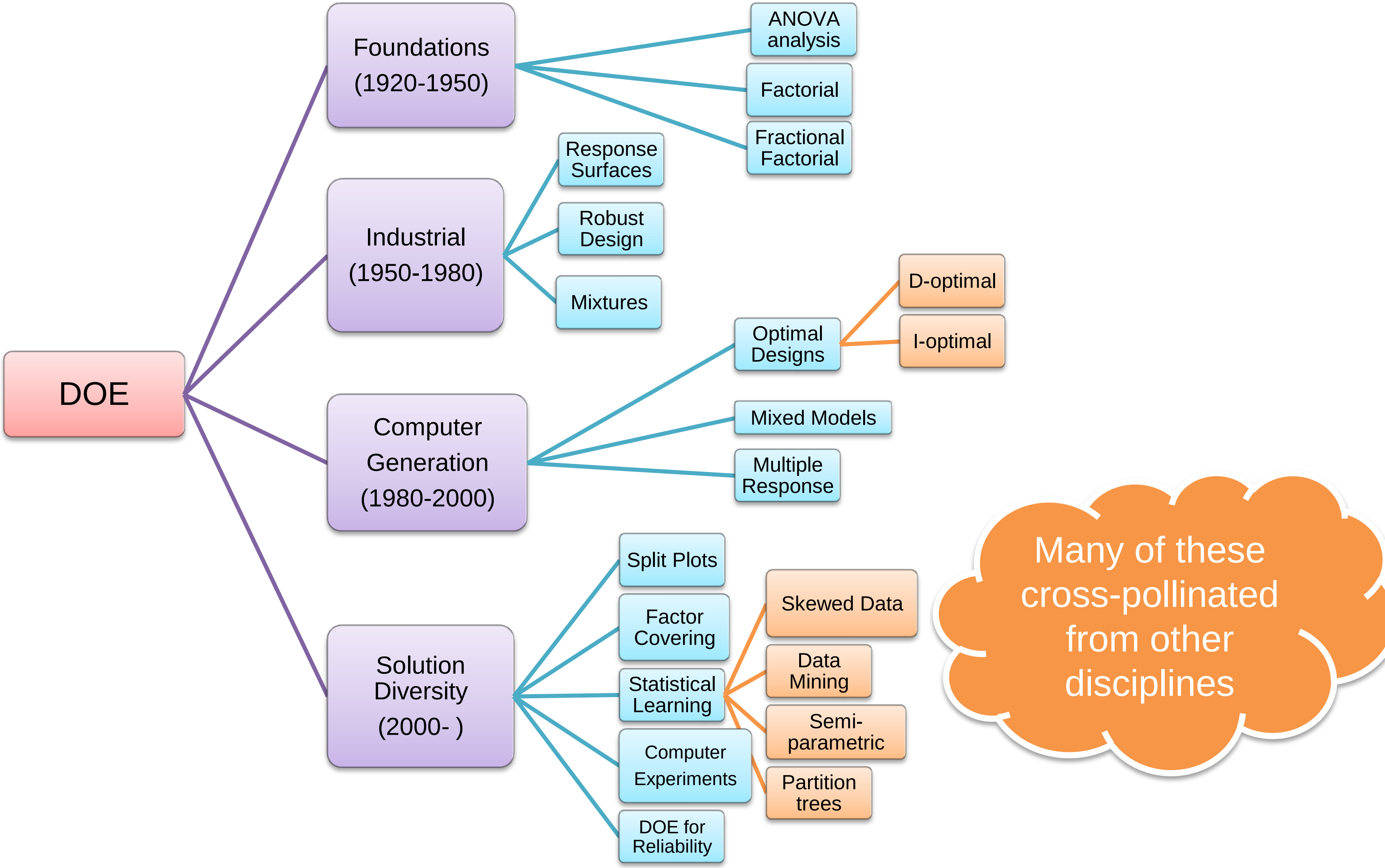
What's in a Name?

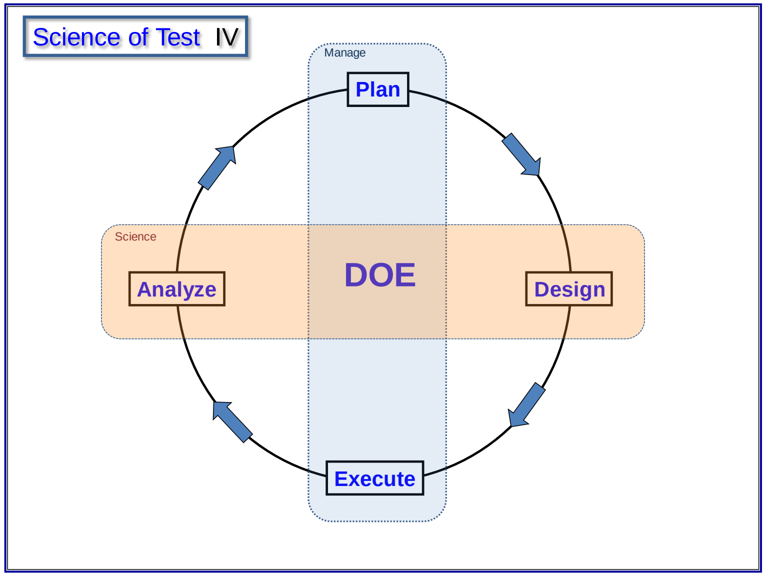
- DOE or even Design of Experiments has downside
  - DOE – Energy, Education …?
  - We already design experiments
  - We test, we don't experiment
  - It isn't just DOE, we need a supporting cast of methods
- Label alternatives
  - Operations Analysis, Industrial Statistics
  - Statistical and Probabilistic DOE
  - Statistically Defensible Test
  - Scientific Test and Evaluation Design
  - Test Science or Science of Test
  - Statistical Engineering or Quality Engineering

One Term for All Test Science

- DOE is used for planning, design, execution and analysis
- DOE uses statistical, probabilistic, and mathematical (including operations research) methods
- DOE encompasses the entire history of design and statistical techniques and methods peer reviewed and demonstrated effective
- DOE is relevant to all types of testing: developmental and operational, deterministic and high-noise systems, for all system complexities
- DOE is *not* the solution for one-shot proof of concept or demonstrations

DOE Evolution





Necessary Tools and Concepts

CV-22 TF Flight Test

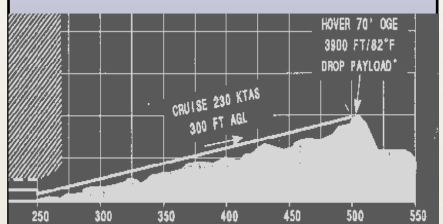
Gross Weight
Radar Measurement

INPUTS
(Factors)

- Airspeed
- Turn Rate
- Set Clearance Plane
- Ride Mode
- Nacelle
- Terrain Type

PROCESS:

TF / TA Radar Performance



OUTPUTS
(Responses)

- Set Clx Plane Deviation
- Crossing Angle
- Pilot Rating

Noise

$$\text{Responses} = f\left(\text{Factors}\right) + \varepsilon$$

Consider the possible effects of three variables: Airspeed, Turn Rate, and Ride

Risks (α and β) Reviewed

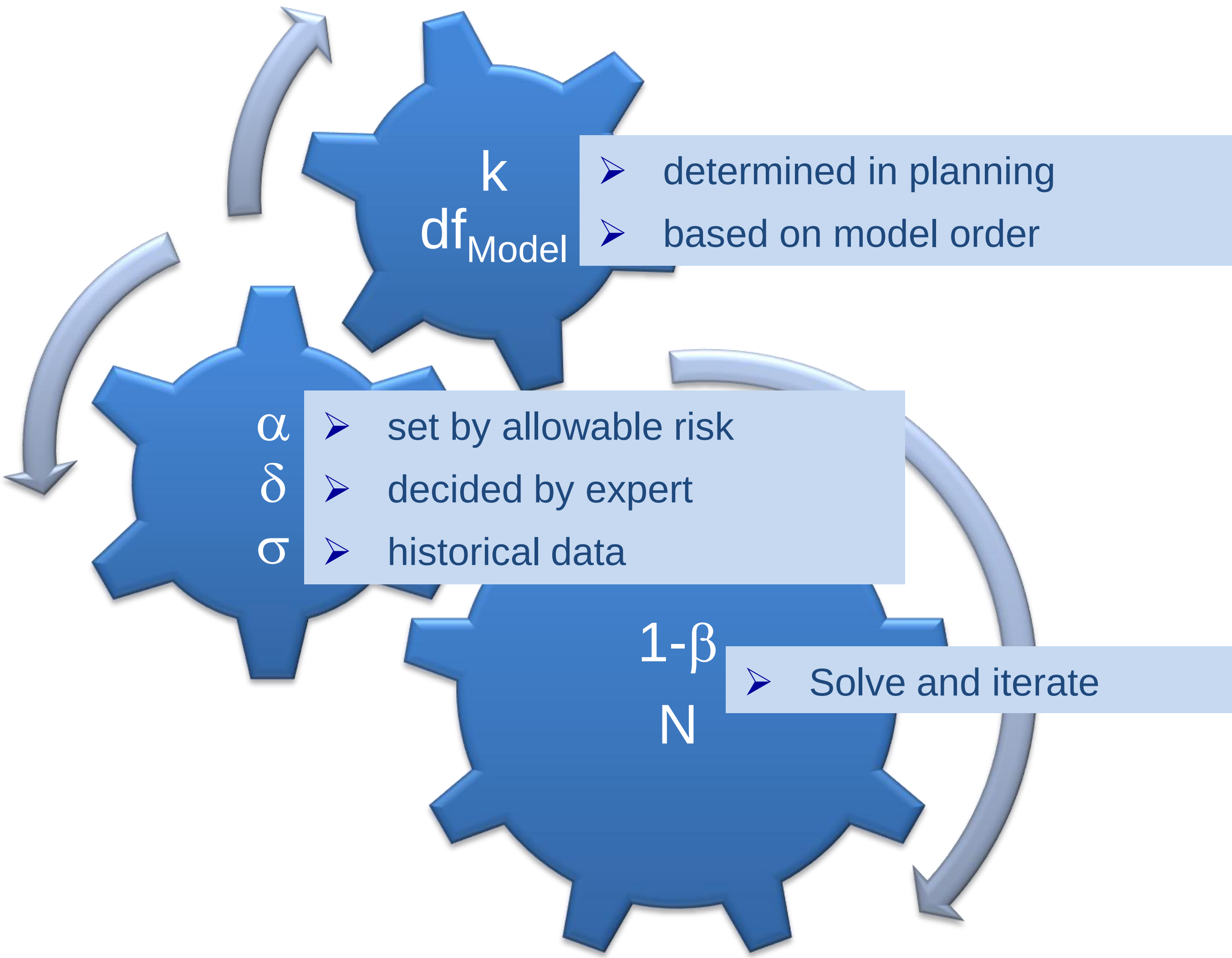
Truth Model: **Response = Ride + Turn**

Test Factors	Hypotheses	Possible Conclusion	Error
A: Airspeed	**H_0: Airspeed has no effect** H_1: Airspeed matters	Airspeed matters	α
B: Turn	H_0: Turn has no effect **H_1: Turn matters**	**Turn matters**	**None, $1-\beta$**
C: Ride	H_0: Ride has no effect **H_1: Ride matters**	Ride has no effect	β

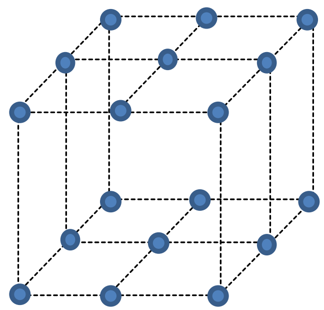
*** Bold Blue reflects the truth**

23

Power Analysis Sequence

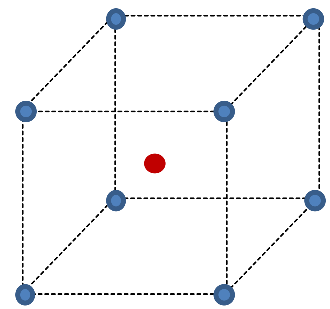


Classic Experimental Designs



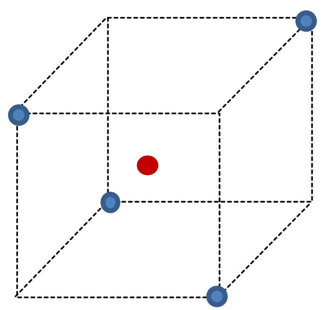
General Factorial

3x3x2 design



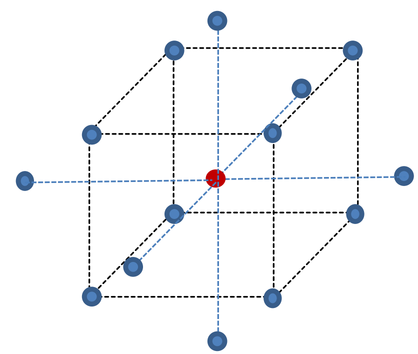
2-level Factorial

2^3 design



Fractional Factorial

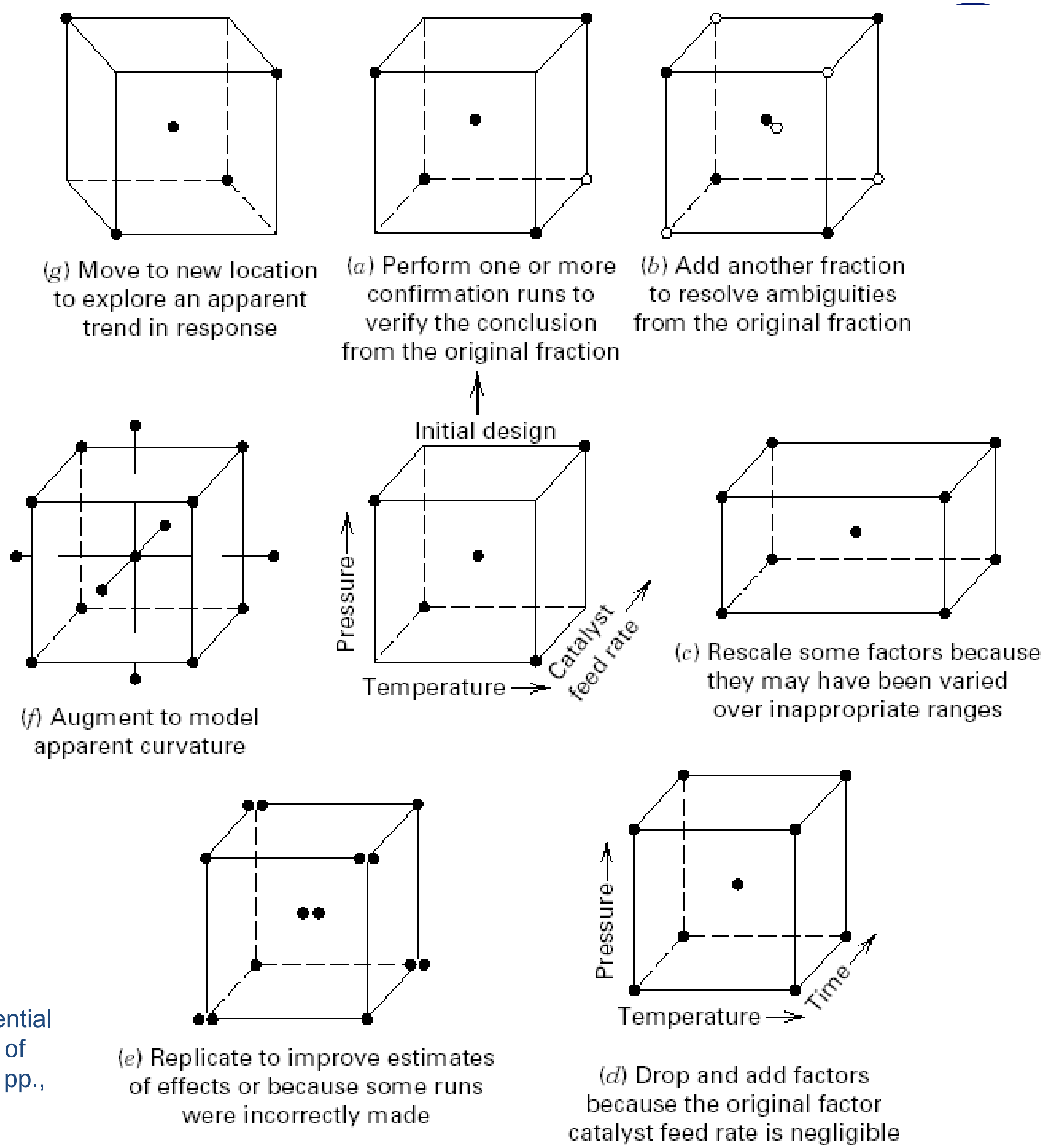
2^{3-1} design



Response Surface

Central Composite design

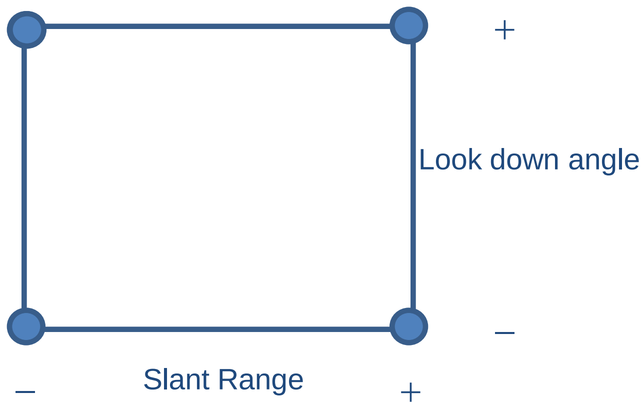
Possible Strategies for Follow-Up Experimentation Following a Fractional Factorial Design

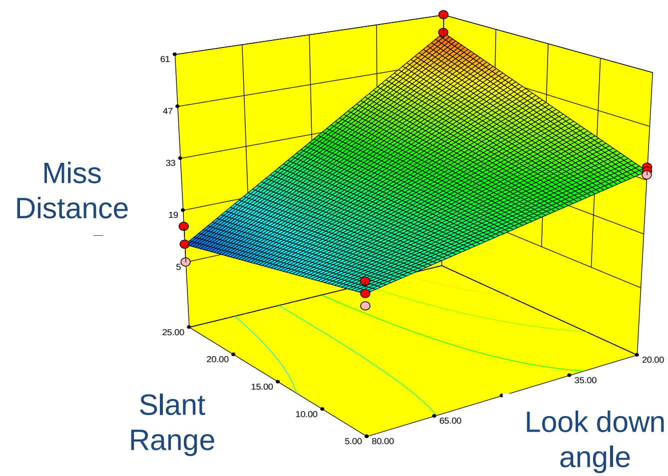


Adapted from Box, GEP (1992-1993), "Sequential Experimentation and Sequential Assembly of Designs," *Quality Engineering*, Vol 5., No. 2, pp., 321-330.

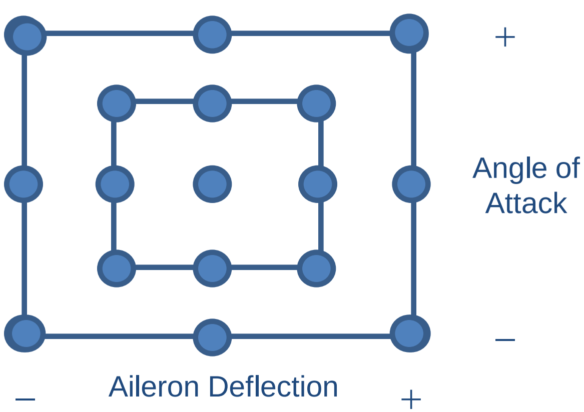
Designs Support the Model

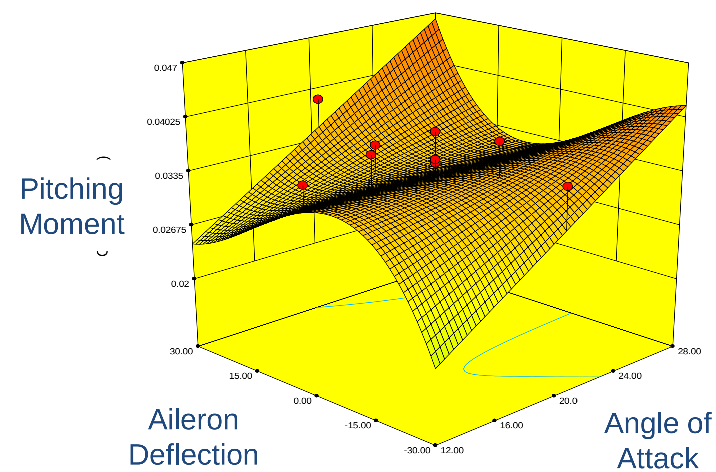
a)





b)





Standard Modeling
Least Squares Regression

Linear in parameters

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

Quantitative
Continuous

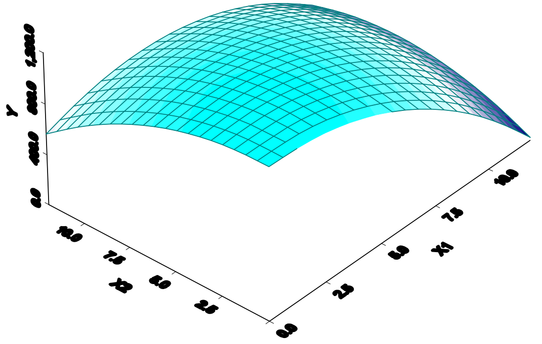
Quantitative
Continuous

Normally distributed
Independent
Homogeneous variance
Single error component

Low correlation

Run	A	B	C
1			
2			
3			
4			
5			

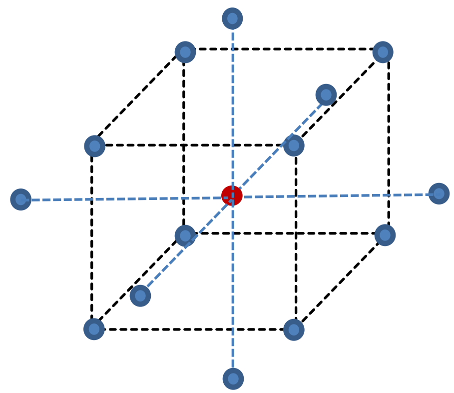
Void of
outliers, leverage points



2^{nd} Order Designs

Assumptions
Randomized
Numeric or Categorical
Mostly Numeric
> 2 level

Design



Attributes
Replication
2^{nd} order design
Nearly Orthogonal
Target Prediction and Coefficient Variance
Efficient runs for k < 7

Assumptions
Errors NID (0, σ^2)
Model is adequate
Y well behaved

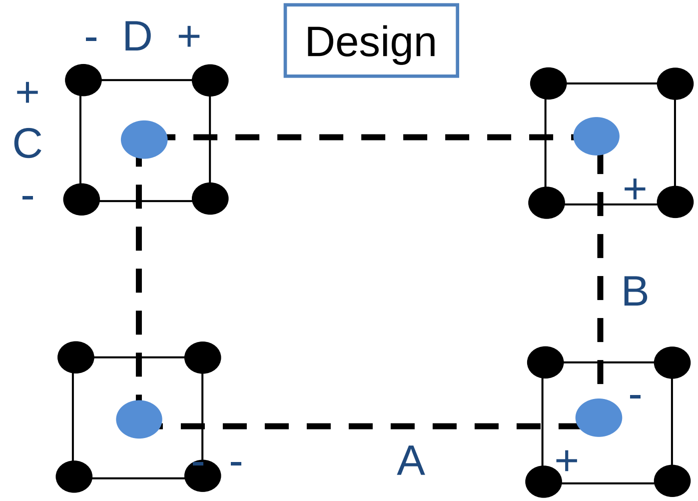
Model

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon$$

Attributes
All effects for general model
Pure error + LOF
Nearly Independent β estimates

Split-Plot Designs

Assumptions
Hard to Change Factors
Numeric or Categorical



Attributes
Replication
Orthogonal

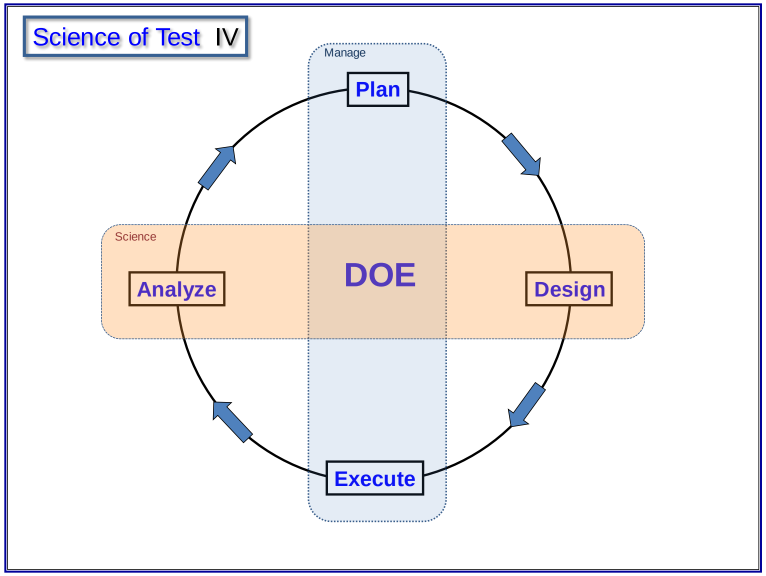
Assumptions
Two Independent Error Terms, both NID (0, σ^2)
Model is adequate
Y well behaved

Model

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \delta + \varepsilon$$

WP error

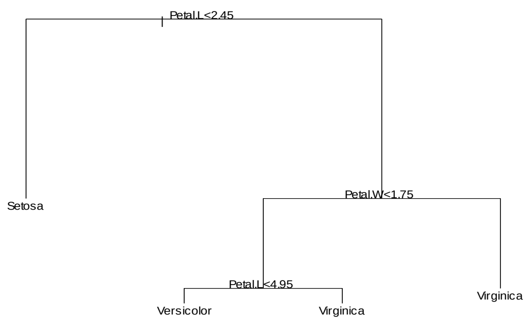
Attributes
All effects of interest
Limited WP error df
Independent β estimates



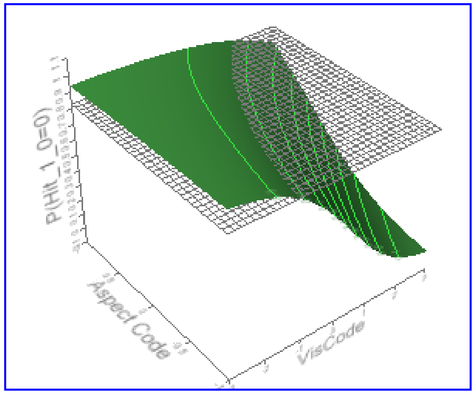
Other Methods

Modeling Alternatives

Tree-based Methods



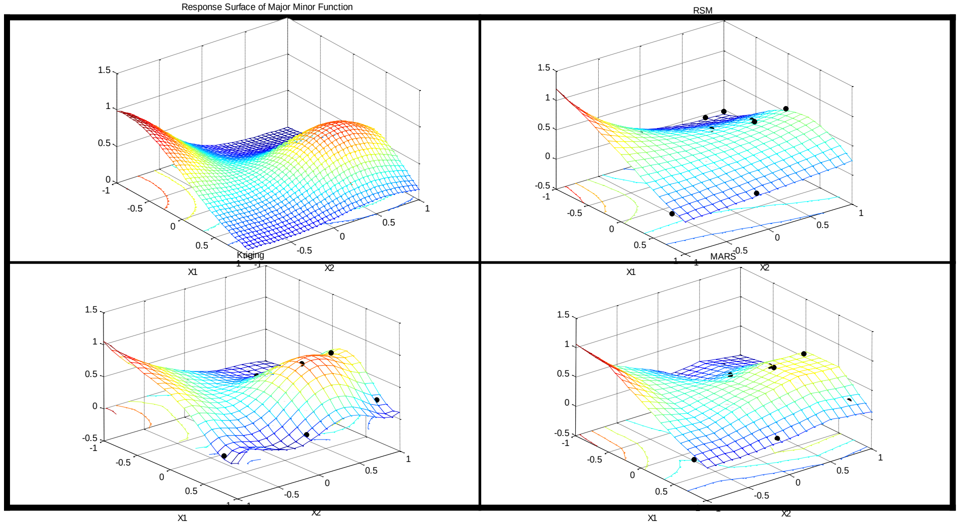
Generalized Linear Models



Nonlinear Modeling

Truth

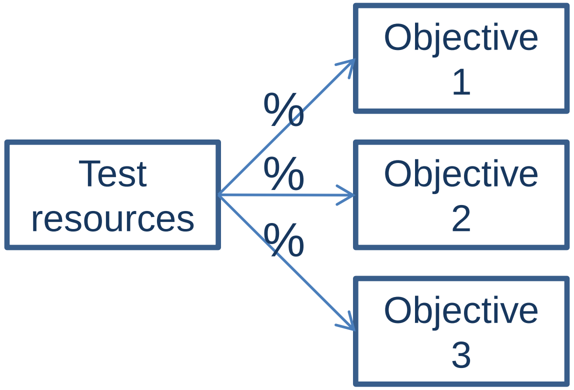
RSM



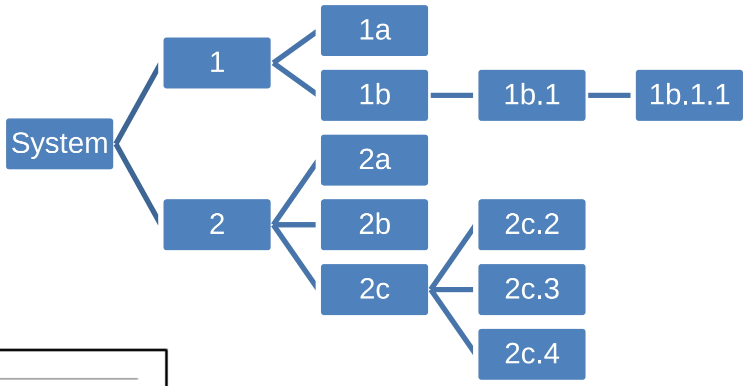
Kriging

MARS

Software Testing Solutions

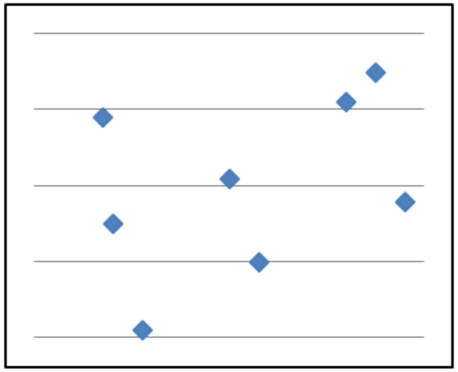


Decision Analysis



Factor Covering Arrays

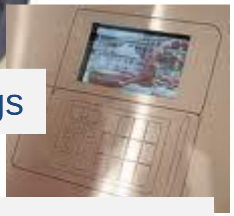
Space Filling



- How to spread out test resources effectively/efficiently
- How to test configurations effectively/efficiently
- How to fill a space effectively/efficiently

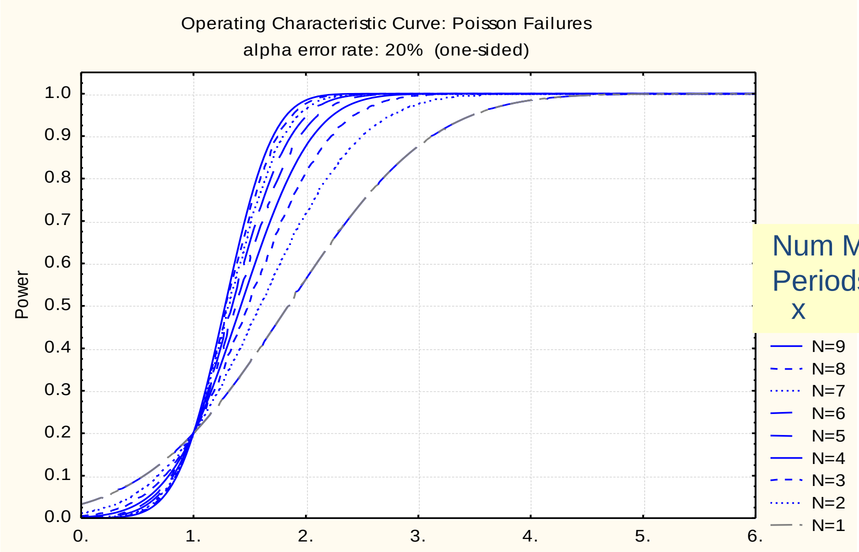
Reliability

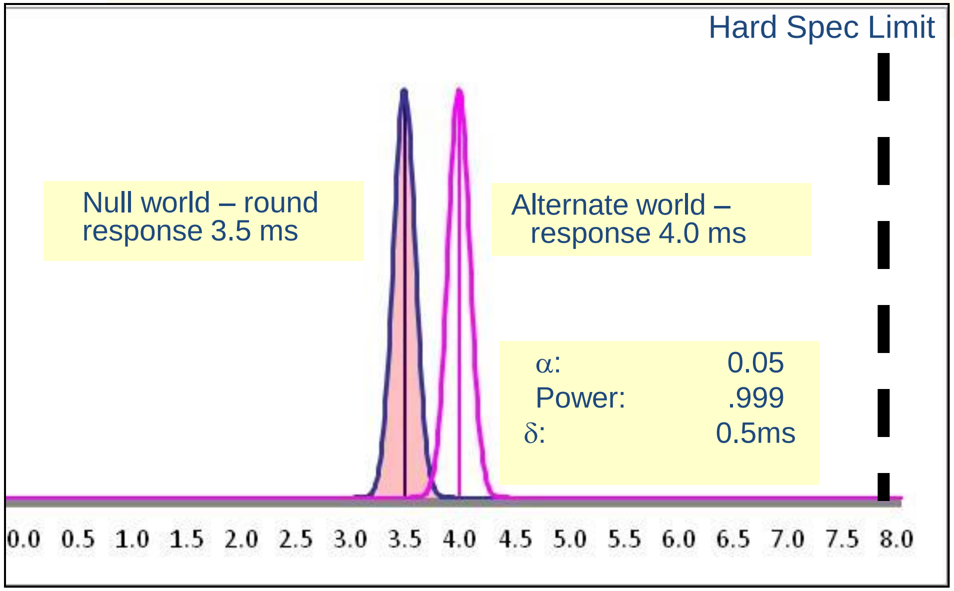
Ceramic bearings



LCD screens

GPS/INS Unit





DOE Mandate Summary

Train	• Training Program • Mentoring – Train the Trainer • Right Methods – Sound & Practical
Practice	• Short Term Wins – Work Projects • Solve Tough Problems • Research and Complement
Lead	• Leadership Commitment • Organizational Adoption • Metrics and Policy

Safety and Reliability for NextGen

System Safety and Reliability Modeling for the Next Generation of Air Transportation

Vitali Volovoi
School of Aerospace Engineering Georgia Tech
E-mail vitali@gatech.edu

NASA Statistical Engineering Symposium

Williamsburg, Virginia

May 5 2011

Risks and hazards of the second order

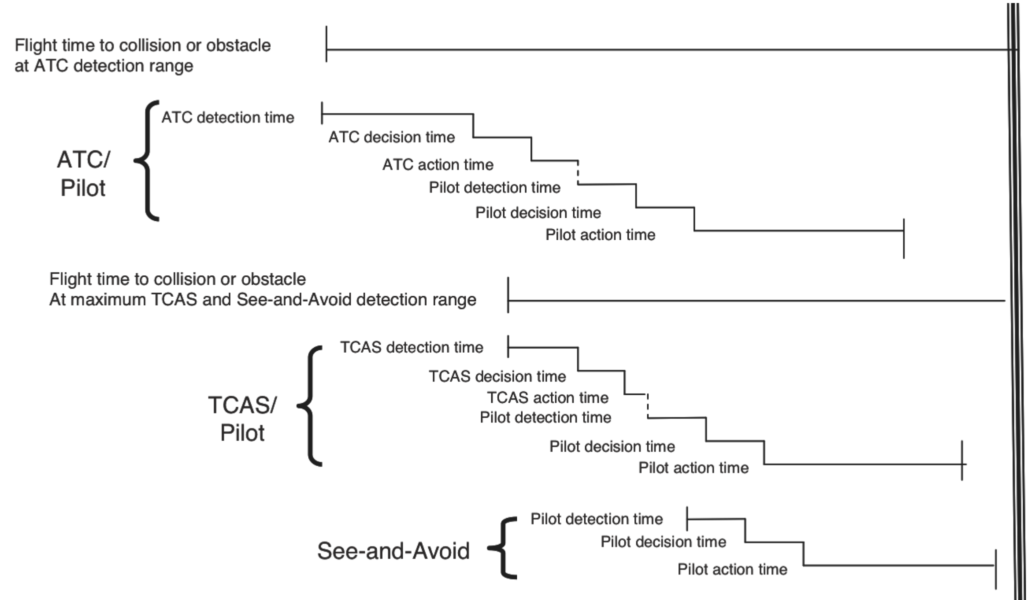
- **Two extremes – the same conclusion:**
  - Existing tools are sufficient and do just fine in modeling safety, so no further research is needed
  - Problems are so complex that there is no point of dealing with them now (we cross the bridge when we come to it)
- **Extensive specific domain knowledge is required as the underlying processes are unique and involved**
- **There is a lack of cohesion in engineering and scientific community making it difficult to ensure that safety issues are adequately addressed**
- **This undercuts the trust from the decision makers (negative feedback loop):**
  - The decision makers do not invest in system safety problems
  - So the community further dissolves

Background

- **Existing approaches to assessing system-level safety:**
  - At the vehicle level: fault-trees and reliability block diagrams FAA certification following ARP4761
  - At the ATC level: combination of fault-trees and event-trees. Similar to Probabilistic Risk Assessment (PRA) used in Nuclear and NASA Space program
- **Both of those existing approaches decouple temporal (event trees or their equivalent) from logical complexity (fault trees or their equivalent)**
- **ATC operations exhibit highly coupled behavior between the temporal and logical domain, and this coupled behavior must be modeled at the bottom level using physics-based simulation**
- **There are two issues with modeling coupled failure behavior at the bottom level:**
  - Breadth vs. depth trade-off in complexity – those simulations are really good depth-wise, not so good from the breadth viewpoint
  - Focus is on the simulation of the operation, rather than on paths and logic of failure propagation

Coupling of timing and logic complexity

- **Timing is important! Static tools (fault trees) cannot capture the timing effects**
- **Simulation with 4-D trajectories is possible, but not practical as the control logic gets more and more complicated (the breadth issue)**

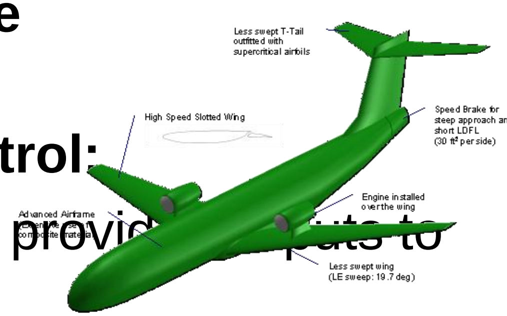


Redundancy of space conflict resolution in current NAS (R. Hemm and A. Busick, ATIO 2009)

- **Stochastic Petri Nets (SPNs) are suggested as an intermediate layer of analysis. Nested analysis is modular (unlike integrated application of SPN to safety of spacing separation– H. Blom et al, ATIO 2007)**
- **Succinct representation: compact discrete state-space and continuous time (more complex is not always better)**

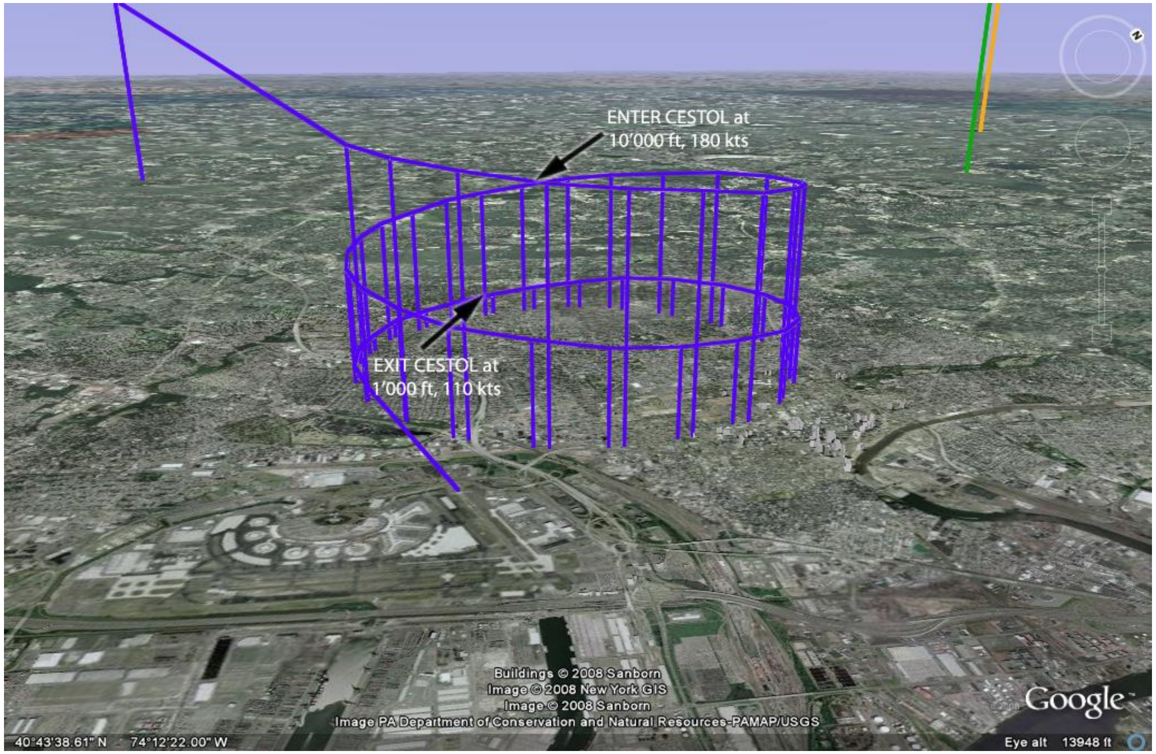
Credible Hazard Scenario for CESTOL

- **Cruise efficient short takeoff and landing CESTOL**
- **Spiral/Helix approach**
- **Impact of wind (steady-state and gust) on the trajectory in the regime of manual control**
- **Potential triggers of reverting to manual control:**
  - Generator/electrical failure of the equipment provid[illegible]puts to FMS
  - Loss of navigational inputs to FMS or degraded state of FMS itself
- **Motivation (potentially exacerbating factors)**
  - Changing heading (changing relative influence of the wind)
  - CESTOL low wing loading – more susceptible to wind disturbances



Spiral/Helix

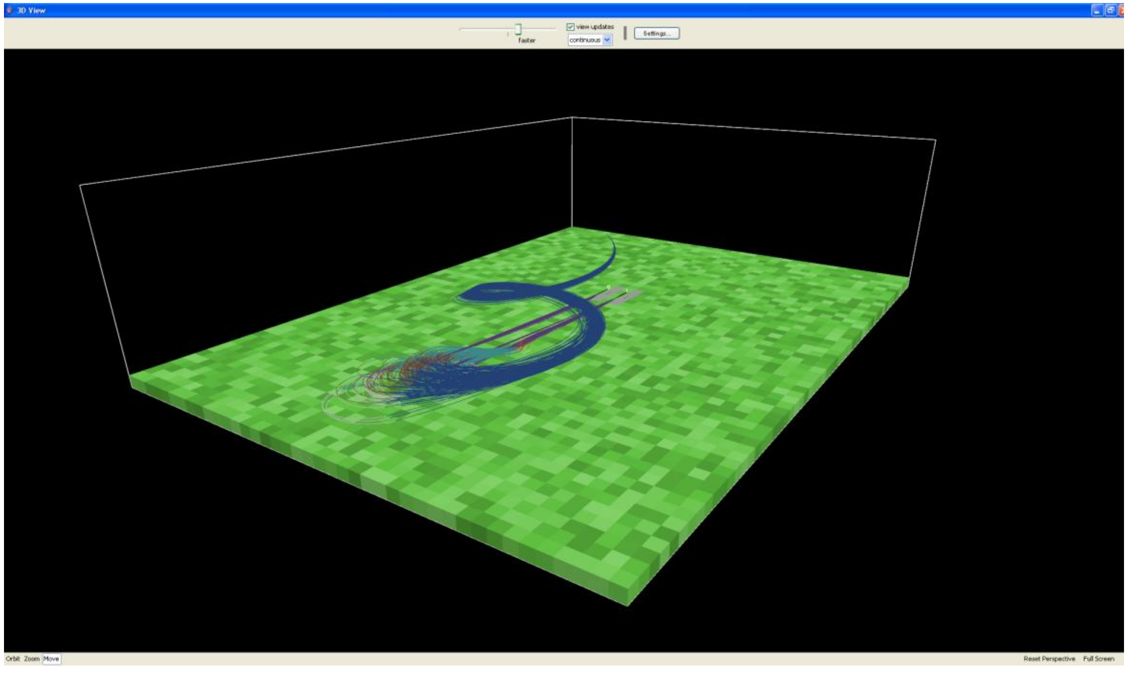
- Spiral descent originally used as noise abatement
- Moved off-airport to allow for stabilization maneuver
- Bank angles was allowed to vary and instead the radius is kept constant (helix)



Descent Helix for CESTOL (Image of Air Transportation Lab at Georgia Tech)

Agent-Based Simulation of using NetLogo

- NetLogo has been developed at Northwestern University, has good interface with other pre- and post-processing software (Matlab, Mathematica)
- Main Parameters: geometry, wind, pilot response delay



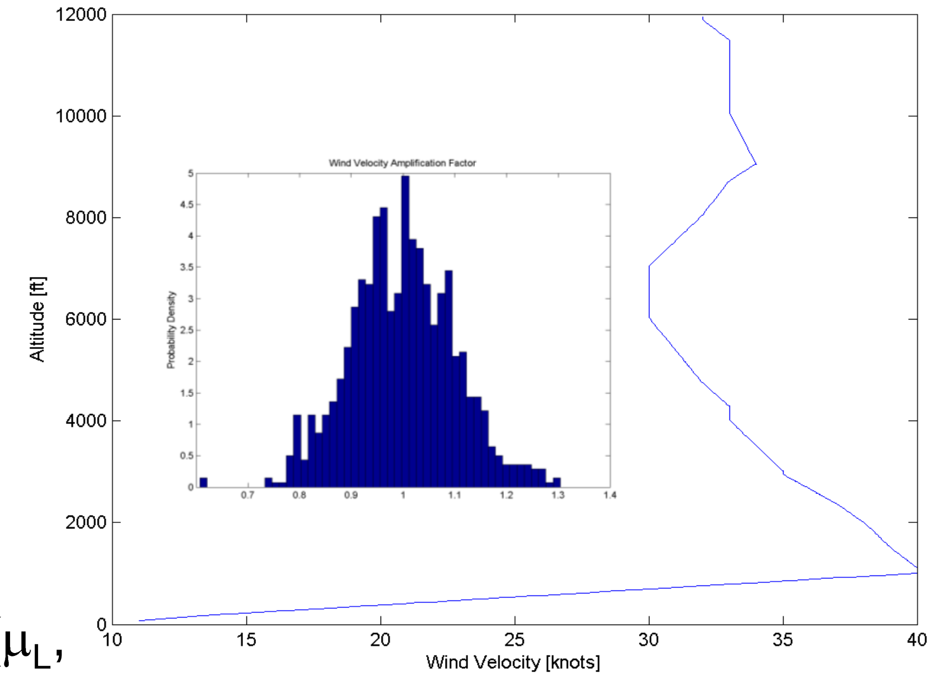
NetLogo model 500 random trajectories

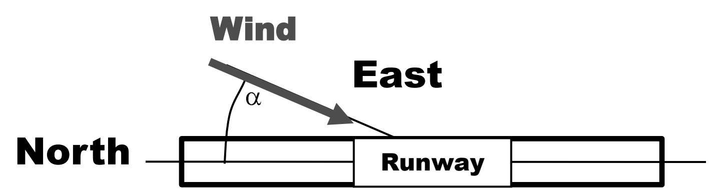
Simulation assumptions

- **Uncertainty Sources:**
  - Pilot's response delay
  - Forecast error (wind direction and intensity)
- **Assumptions:**
  - Rescaling & rotation of nominal wind according to Gaussian distribution ~$N(\mu, \sigma)$
  - Lognormal pilot's response ~ $L(\mu_L, \sigma_L)$ (not to exceed 20 sec)
- **Probability of minimum distance violation**

Focus of the model

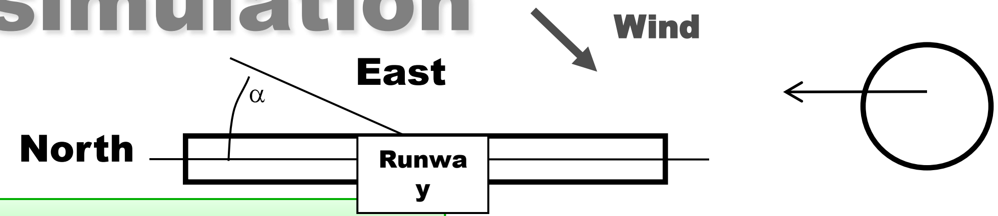
$$P_{violation} = P_{failure\_fms} \times P_{helix\_drift} \times P_{other\_aicraft}$$

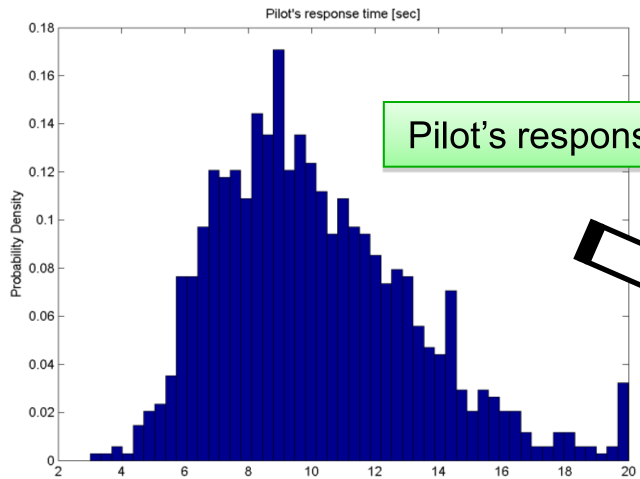




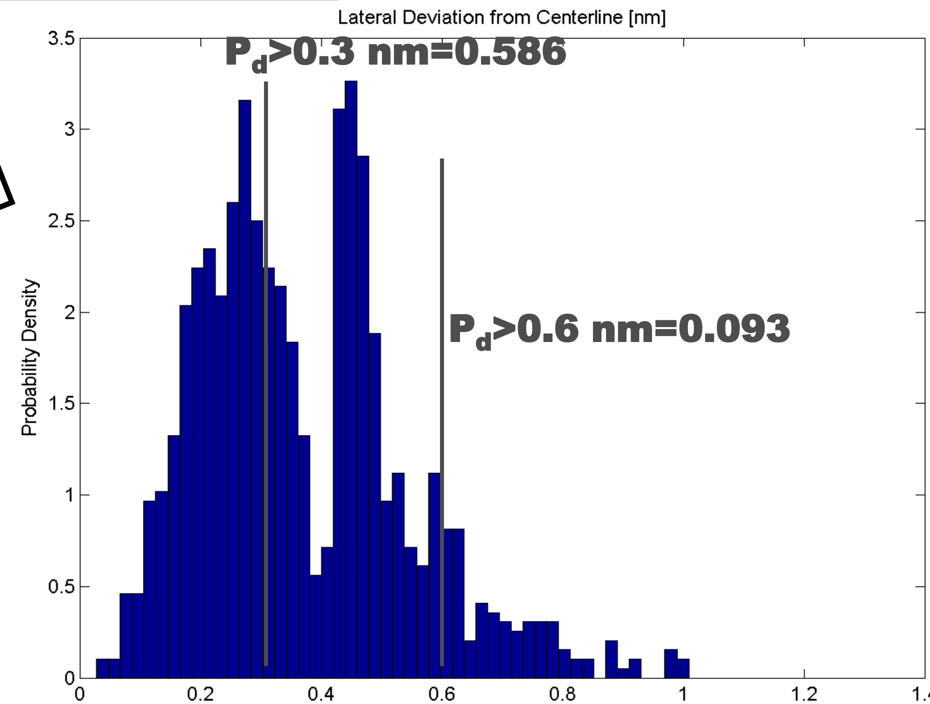
In general, things are a bit more complicated (timing is important)

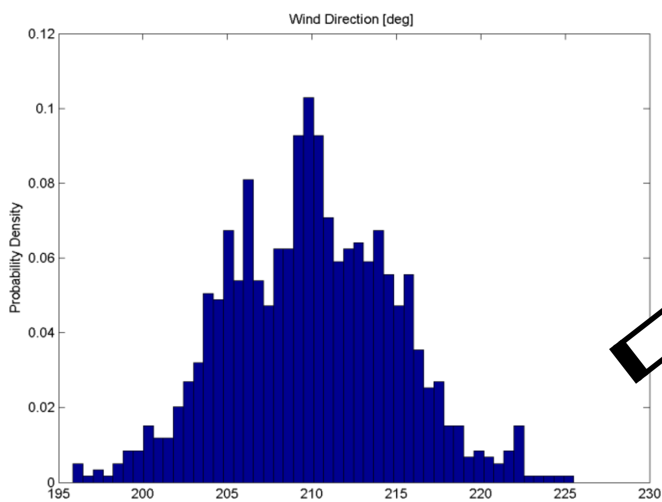
Results from agent-based simulation





Pilot's response time ~ L(μ_L=10 sec, σ_L=10 sec)

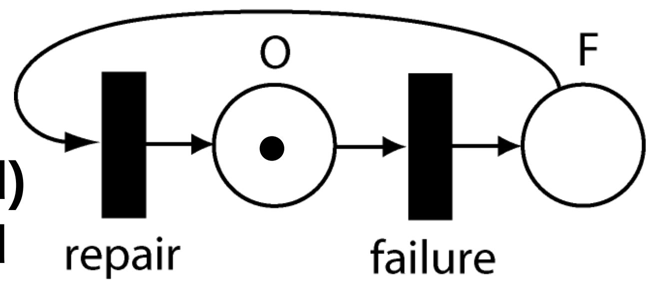




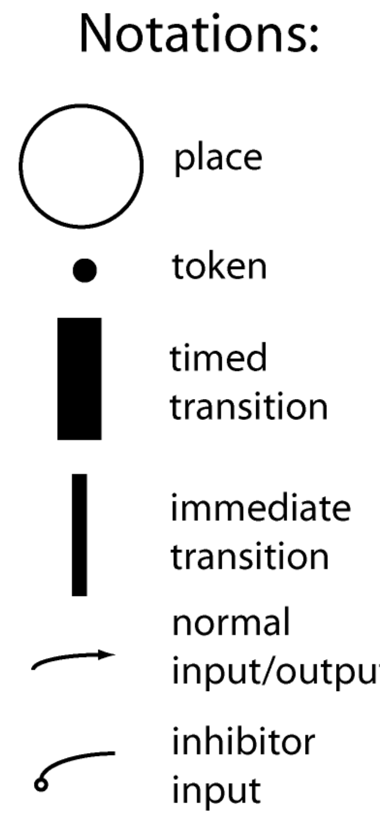
Wind intensity's amplification factor ~ N(μ=1, σ = 0.1)
Error in wind direction ~ N(μ=0°, σ= 5°)

Introducing Stochastic Petri Nets

- Tokens represent relevant entities of a modeled system
- Places represent possible states of those entities
- Tokens occupy places, thus realizing particular states of the corresponding entities
- The combination of all tokens' locations (so-called marking) uniquely characterizes the modeled system
- Tokens move between places, simulating changes in the system state
- Transitions describe the rules for token movements: tokens are "fired" from one place to another via transitions
- Transitions fire only when they are enabled (i.e., if certain conditions are satisfied)
- Transitions are enabled based on where other tokens are thus capturing interdependence among components states (inhibitors are used)
- An enabled transition fires after a specified delay (the transition's attribute)

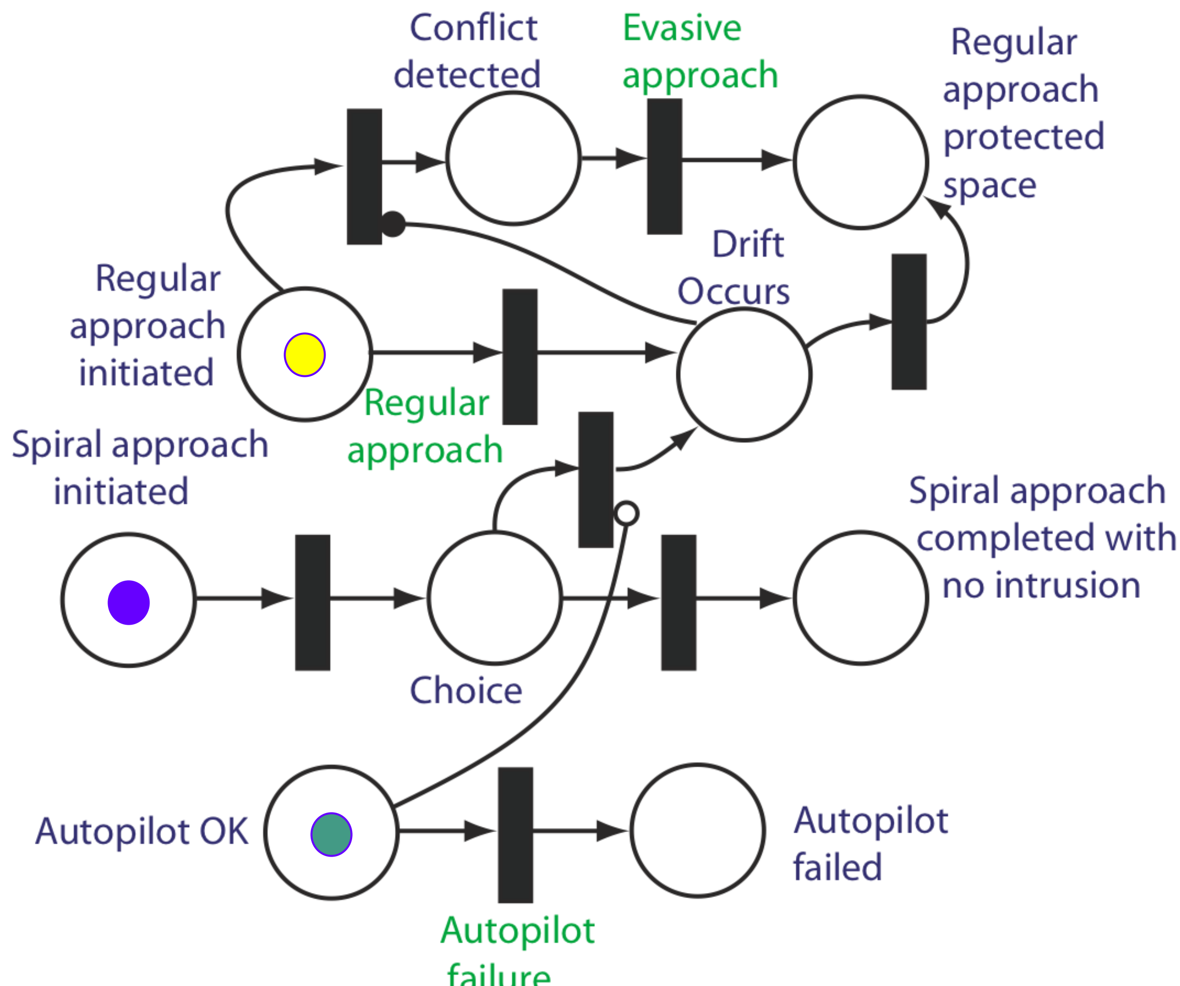


A repairable unit



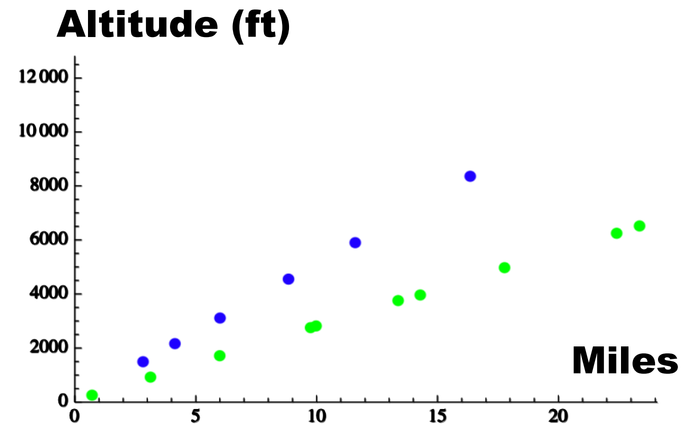
SPN@: Implementation of SPN with aging tokens

Modeling of separation violation at a higher level of abstraction using SPN



A Hazard Scenario from VLJ

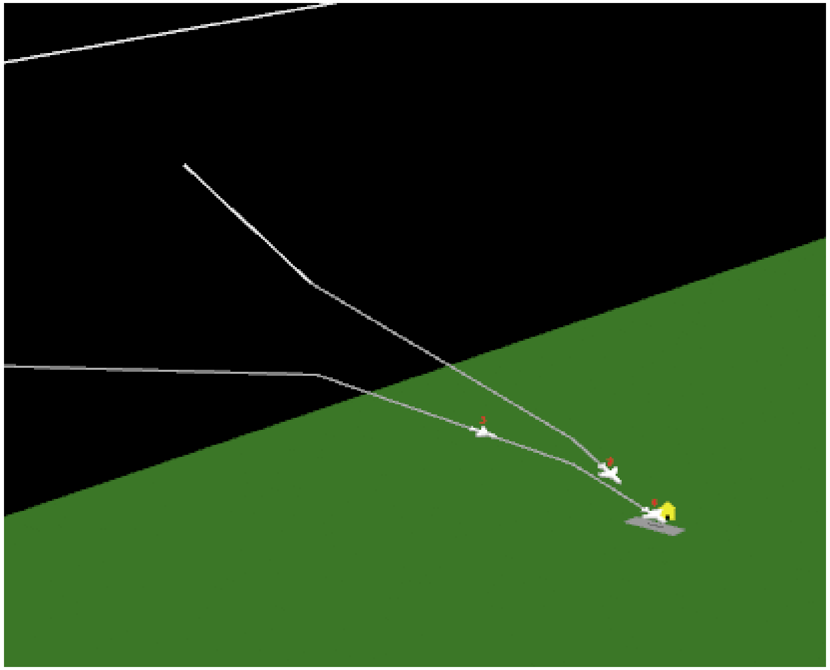
- Very Light Jet on a straight steep (5.5 degree) approach mixed with two regular aircraft loses capability to evaluate the altitude (e.g., Pitot tube obstruction resulting in corruption of air data) and starts to descent faster than intended thus potentially leading to vertical space violation with the leading aircraft



- ATC notices the impending loss of separation and orders leveling off
- If VLJ is not responding after a certain amount of time the leading aircraft is ordered to speed up the descent
- Single pilot vs. two pilots (remote co-pilot)
- Motivation – importance of accommodating mixed approach with various descent speed by means of vertical separation
- Motivation – importance of investigating the viability of a back-up pilot on the ground

Challenges of modeling VLJ

- Standard load-sharing by pilots (where in the case of emergency one pilot flies and the other trouble-shoots the problem and communicates with ATC) is not applicable when one of the pilots is on the ground
- As a result, the co-pilot on the ground is assumed to have the same capabilities as the VLJ pilot



Snap Shot of NetLogo model of VLJ

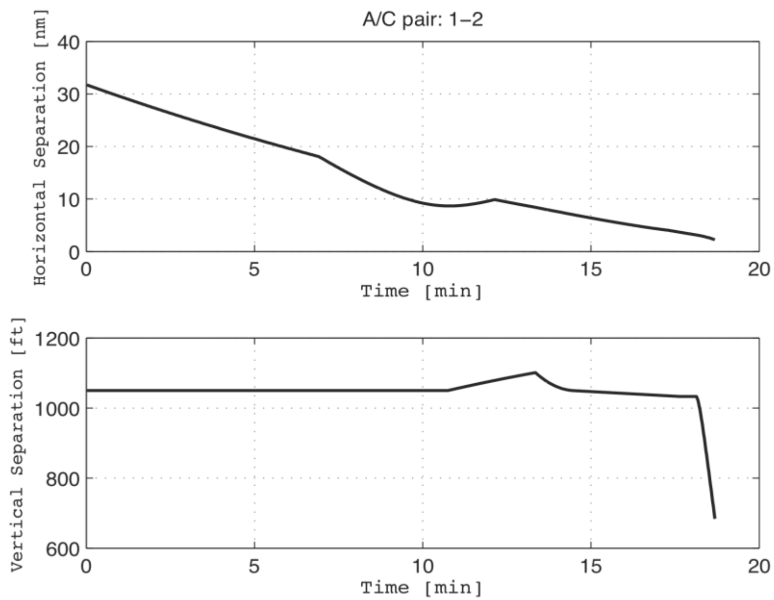
- We assume that pilots share the load, thus conducting tasks faster (up to twice as fast). When ATC sends the command, the current task is completed, and then ATC command is executed

Pilot Tasks breakdown and their duration

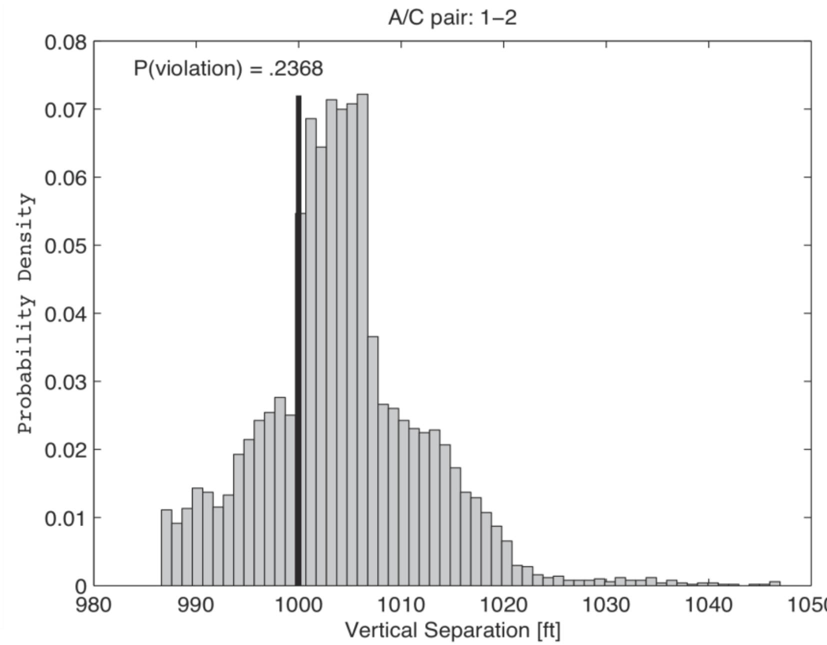
Phase of flight	Task	Average time to accomplish (mins)	STD
Prior to Top of Descent	Check Weather	2.5	0.65
	Check fuel state	0.5	0.158
	Check altimeter setting	0.5	0.149
	Check FMS programming	1	0.3
	Select radio frequencies	0.5	0.176
	Review Approach Plate	1	0.25
	Receive Approach Clearance [ATC communication]	varies depending on clearance	
	Set DH or MDA into Radar Altimeter	0.25	0.017
	Select landing gear down	instantaneous	
Top of Descent	Monitor Descent Progress	throughout descent (every 1 min)	
	Instrument scan	throughout approach (every 10 secs)	
	Monitor aircraft systems	throughout flight (every 1 min)	
	Frequency Change	0.25	0.015
	Check RAIM [GPS satellite coverage]	0.25	0.027
	Check aircraft configuration	0.25	0.046
Final Approach Fix	Report FAF [contact ATC]	0.25	0.037
	Scan for runway environment	intermittent from FAF to landing	
	Verify landing clearance [contact ATC]	0.1	0.056
	Wind check [contact ATC]	0.25	0.025
DH/MDA	Verify runway environment IAW (FAR 91.175)	until touchdown	
If runway environment is not insight			
Missed Approach	Execute missed approach instructions	depends on instructions	
	select TO/GA	instantaneous	
	Clean-up aircraft (manually)	2	0.8
	Report missed approach [contact ATC]	0.1	0.015

Sample results for VLJ

- Pair-wise separation is studied when the faults are inserted at different altitudes (1000 cases of Monte Carlo agent based simulation)

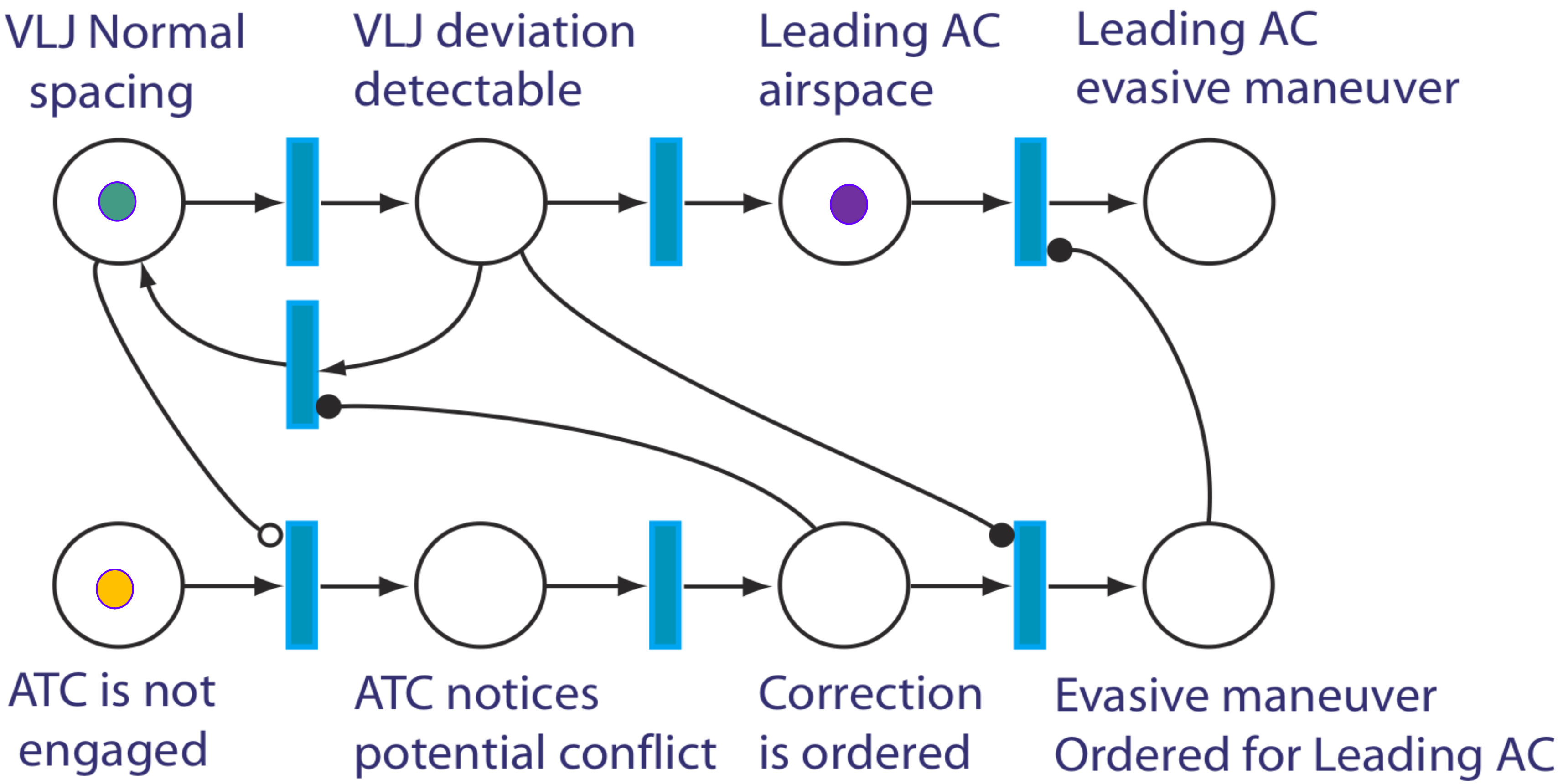


Tracking pair-wise horizontal and vertical separation

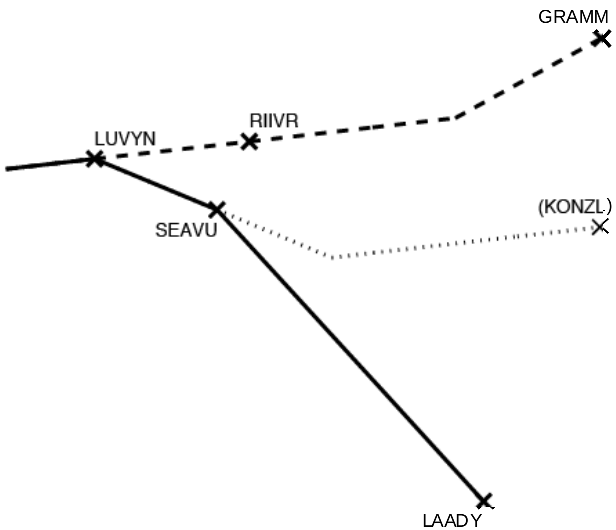


Probability of the loss of vertical separation
1-2 aircraft, loss of VNAV at 5000ft

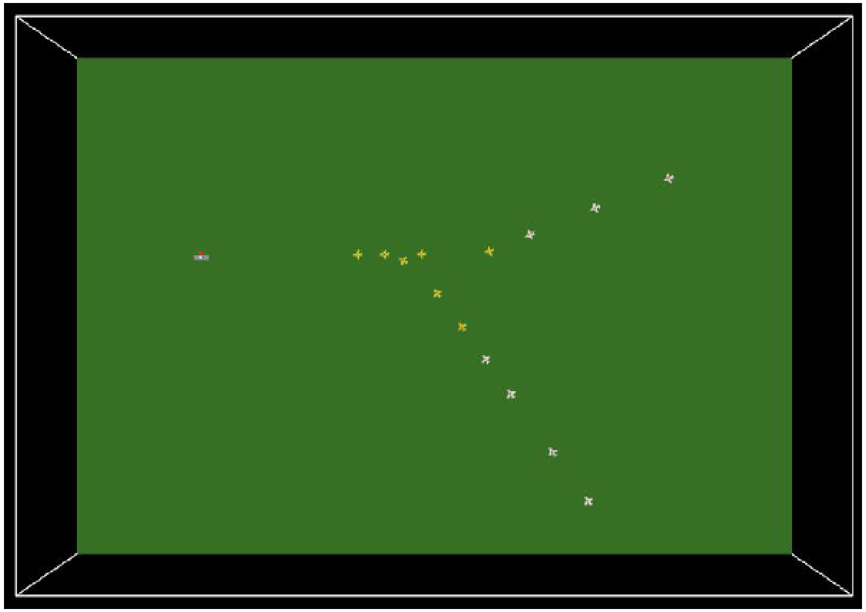
SPN: VLJ Hazard Scenario



Merging aircraft with optimized descent profile in LAX



Schematics of merging routes

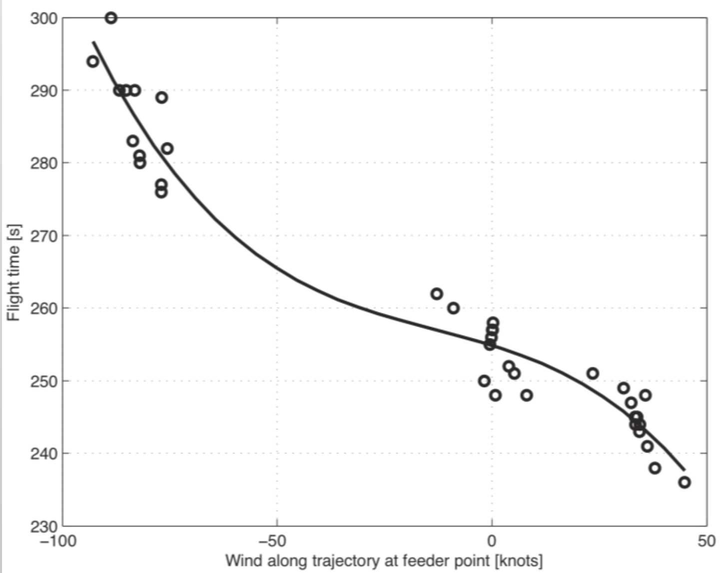


Agent-based simulation (Netlogo snapshot)

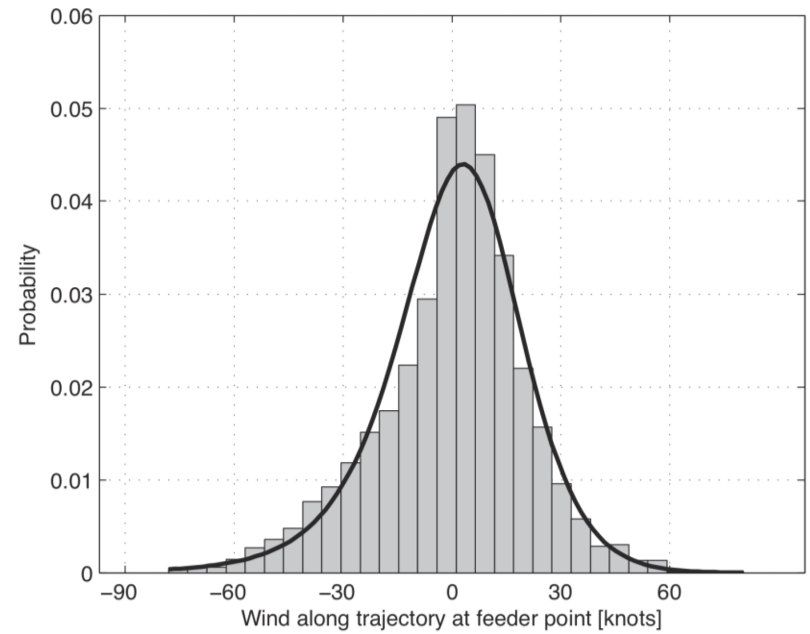
Only a portion of operational procedures in modeled in this example (no vectoring, acceleration or, coordinated conflict is modeled), only two air traffic fluxes are considered

Merging aircraft with optimized descent profile

Optimized descent profile (also referred to as continuous descent) is implemented in LAX – has fuel efficiency and noise benefits but introduces uncertainty in traveling time due to wind

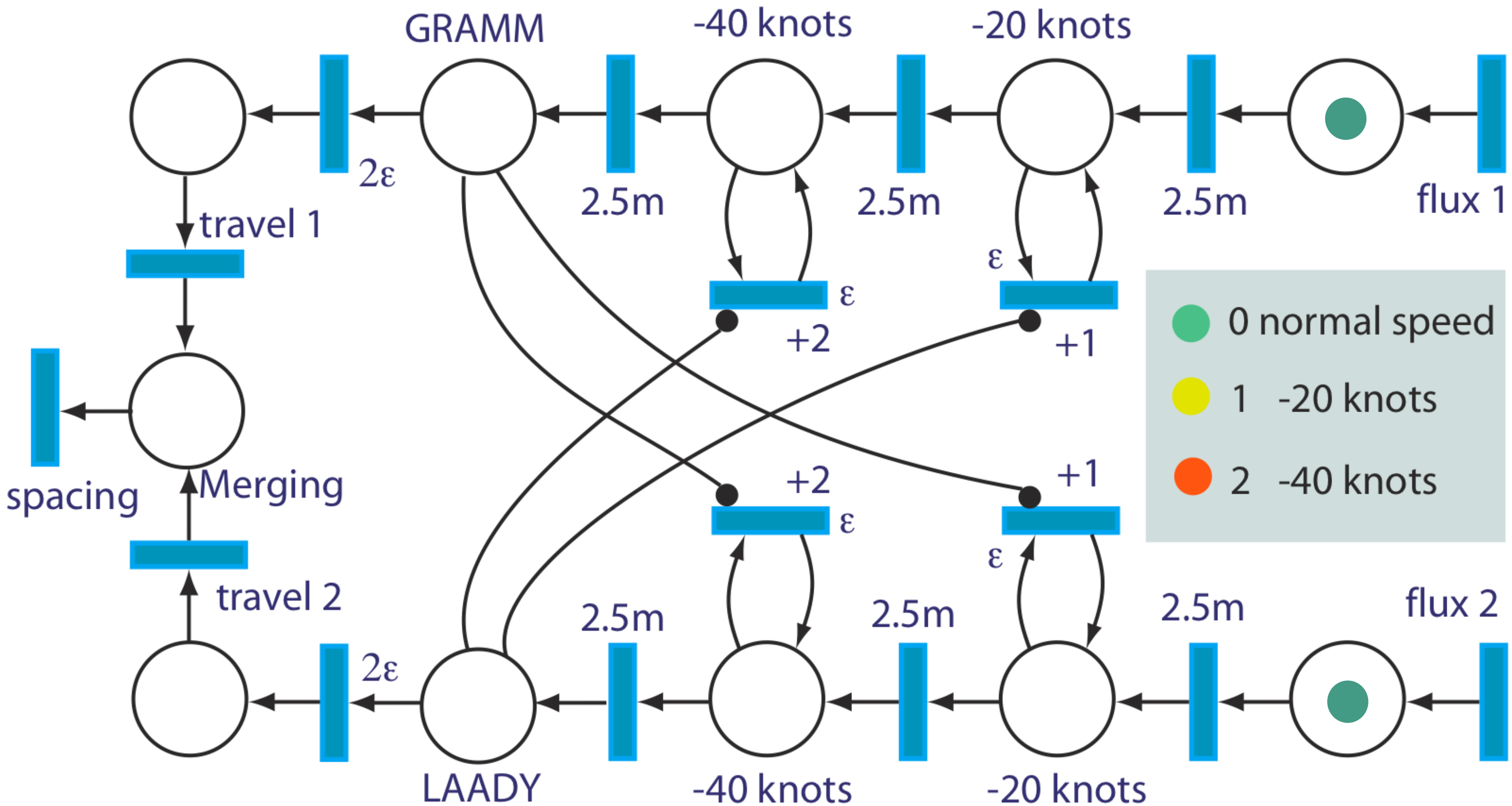


Flight time to merging point as a function of the wind – not that is non-linear!



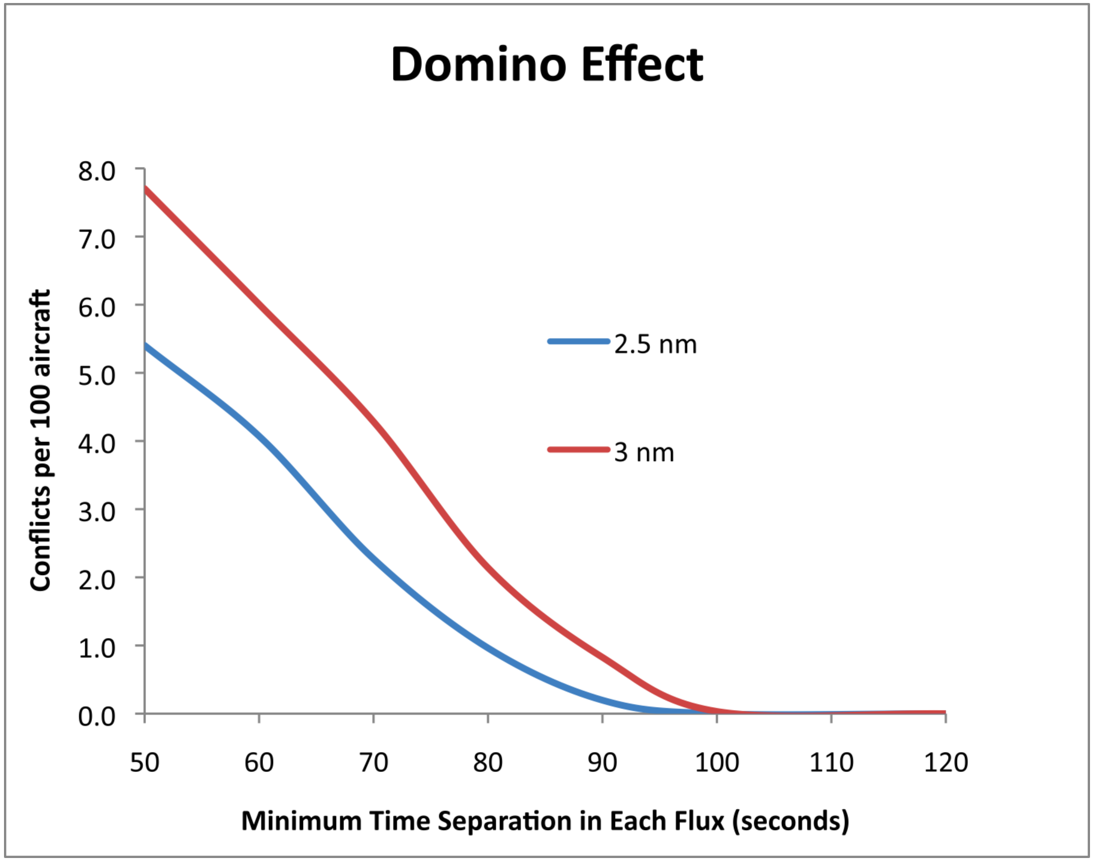
Wind distribution (as observed)

Merging aircraft OPD in LAX – SPN model



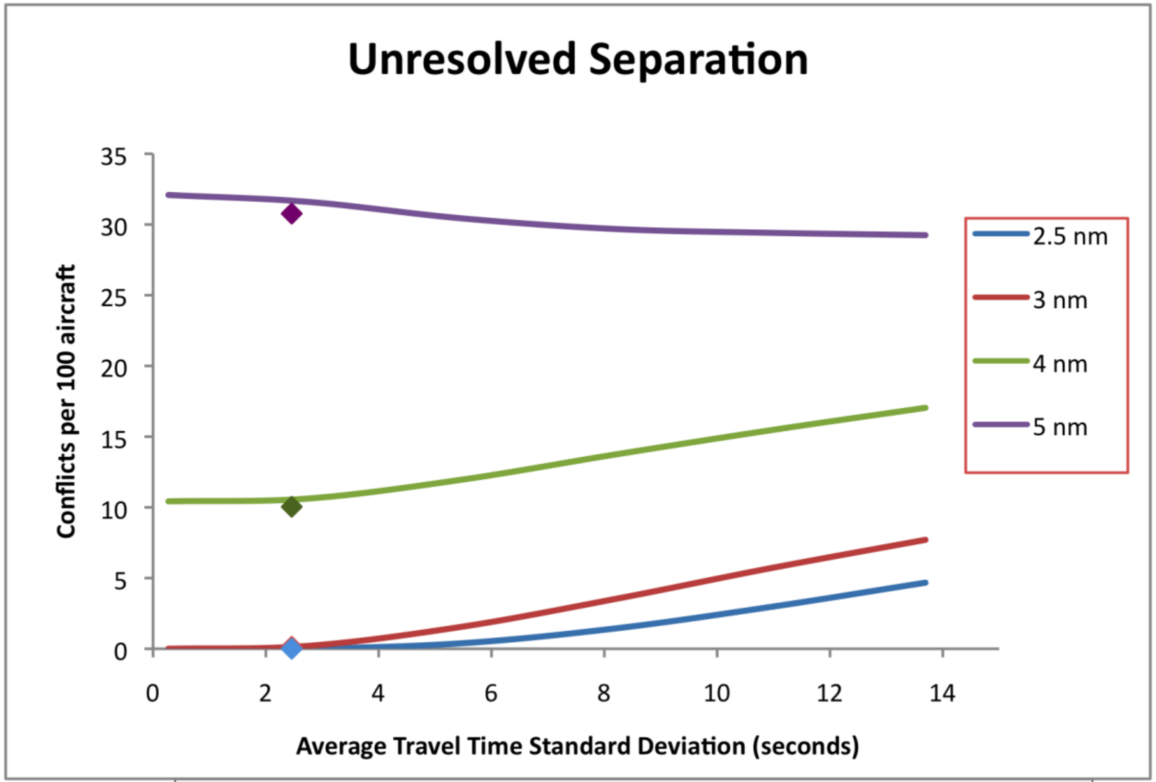
Tokens represent aircraft that change colors in accordance with the ordered maneuvers. Statistics are collected about the conflicts at the merging point (when two tokens are together)

Sample of results from SPN – efficiency of the maneuvers (unresolved conflict frequency)



Frequency of spacing violations as a function of a minimum separation within each flux (no wind is considered)

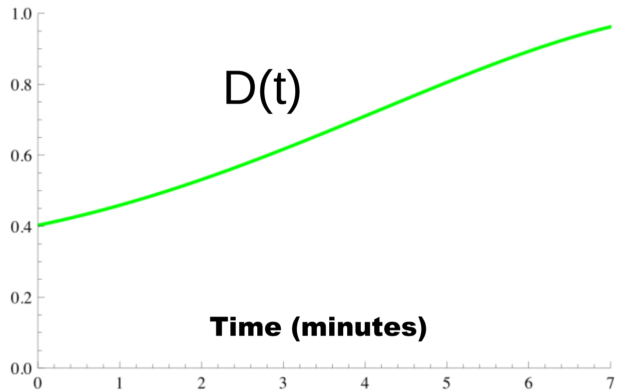
Sample of results from SPN – efficiency of the maneuvers (unresolved conflict frequency)

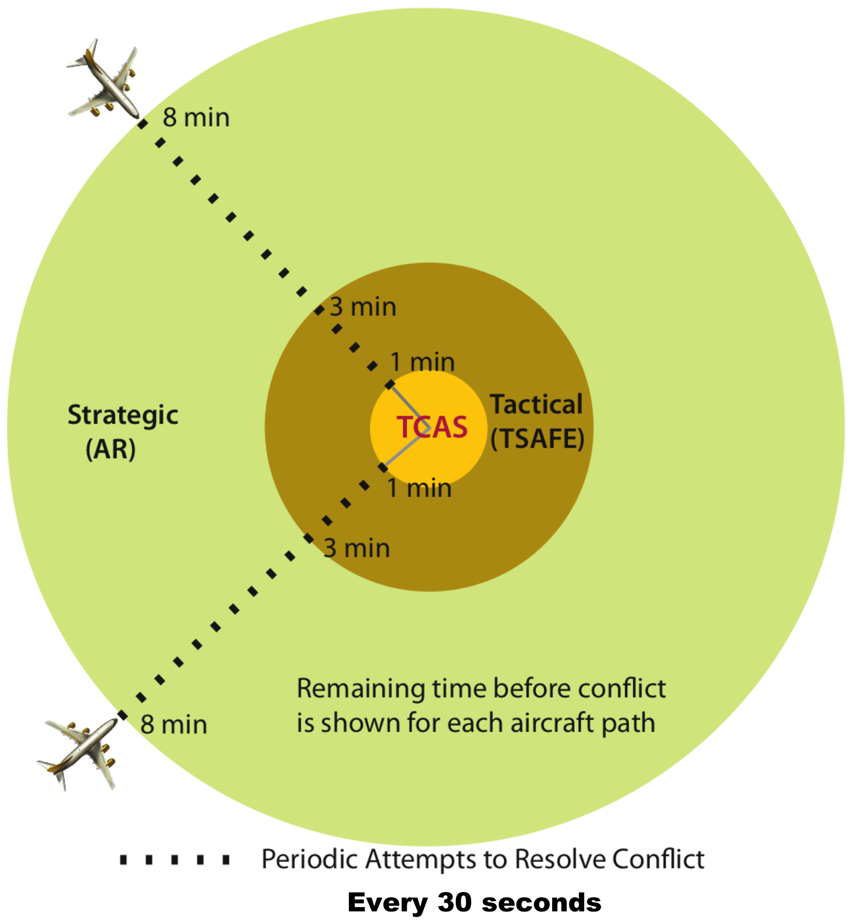


Sensitivity of the maneuver efficiency as a function of the travelling time uncertainty (diamonds represent the results of global agent-based simulation for modeled wind)

Safety Analysis of the Advanced Airspace Concept: State space representation

Probability of Conflict Detection

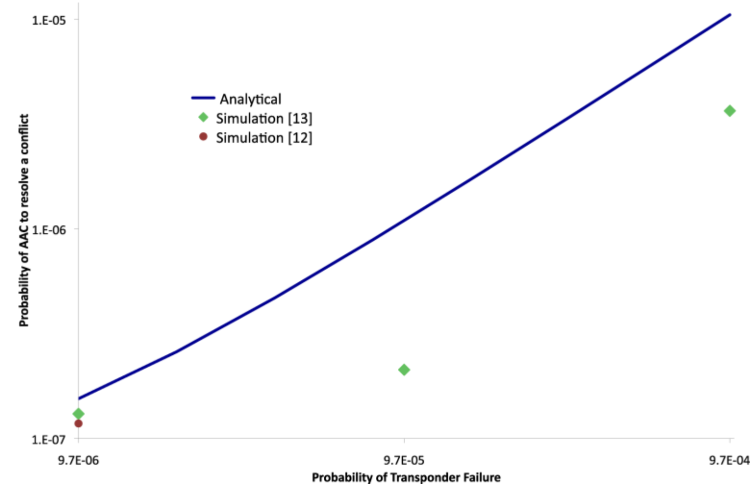




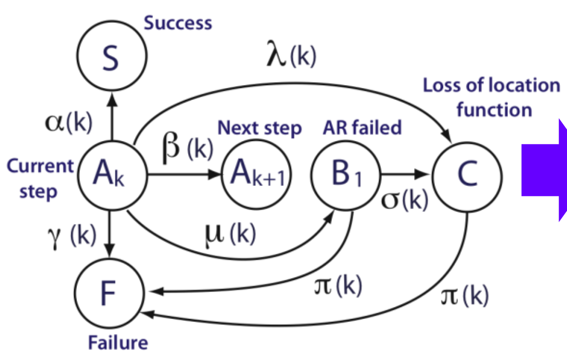
Three layers of automation:
1. Autoresolver (AR) – from 8-20 to 3 min before the conflict
2. Tactical Separation-Assured Flight Environment (TSAFE) 1-3 min before the conflict
3. TCAS – 1 min
+ visual avoidance

Safety Analysis of the Advanced Airspace Concept: State space representation

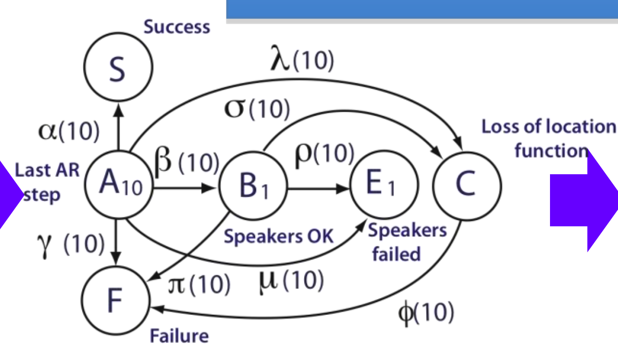
Probability of Sub-system failure is increasing with time, and different layers share common subsystems:
T – transponders (all three layers)
L – location function (AR and TSAFE)
K – speakers TSAFE and TCAS,
in addition subsystems specific to each layer also can fail (A, B, C for AR, TSAFE, and TCAS, respectively



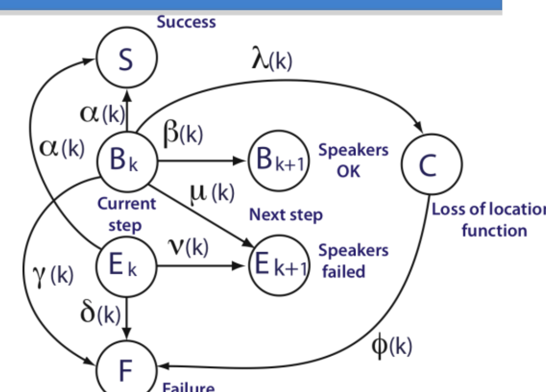
Analytical procedure is developed it Includes modeling of dependent subsystems (by semi-inverting Markov model for non-repairable portions of the system)



AR phase
(time steps 1-9)



AR to TSAFE transition
(time steps 10)



TSAFE phase
(time steps 11-15)

Conclusions

- **Complexities of modeling safety aspects of NextGen should not prevent us from trying our best, as neglecting those aspects will lead to dire consequences**
- **Agent-based simulation provide a useful environment to investigate combined effects of NextGen, procedures, and vehicle characteristics (analogous to physics-of-failure modeling in reliability), but they have their limitations**
- **While realistic logic branching can be modeled using agent-based simulation, a more compact modeling at a higher level of abstraction is beneficial at the very least**
- **Nested hierarchy of models is required for comprehensively assessment of the safety of new vehicle integration into NextGen**
  - Most detailed level: agent-based and simulations of perturbations of 4-D trajectory as well as detailed human-performance models (including Human-in-the-loop simulations of specific scenarios)
  - Intermediate level: Stochastic Petri Nets or analogous discrete-event simulation captures timing event, but provides discrete state-space representation. Markov chains if possible
  - Top level: Fault Tree and similar Boolean Algebra tools

Statistical Design and Analysis of Experiments for Next Generation Air Transportation Research

Sara R. Wilson

NASA Langley Research Center

NASA Statistical Engineering Symposium

May 5, 2011

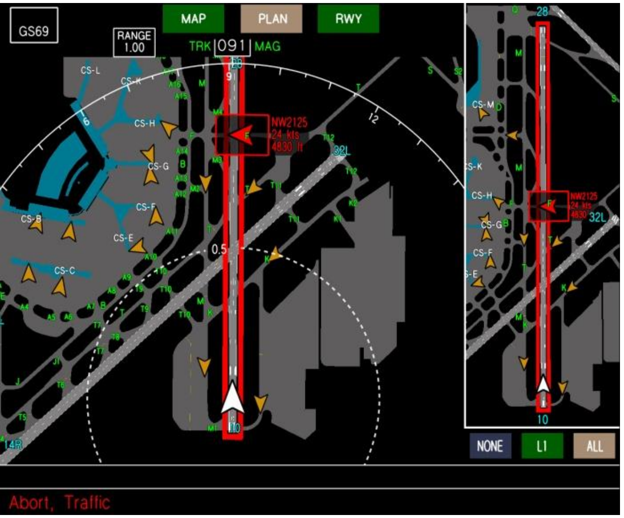
NextGen

- Forecasts project air traffic demand to double by 2030
- The current Air Traffic Management system is already nearing its capacity
- If left unmodified, the current system cannot indefinitely sustain the projected traffic growth without inducing significant delays and inefficiencies
- The objective of the NextGen-Airspace Project is to develop and explore fundamental concepts, capabilities, and technologies to enable significant increases in the capacity, efficiency, and flexibility of the National Airspace System necessary for the Next Generation Air Transportation System (NextGen)
- NextGen Concepts and Technology Development Project
  - Separation Assurance
  - Safe and Efficient Surface Operations
  - Super Density Operations

Safe and Efficient Surface Operations

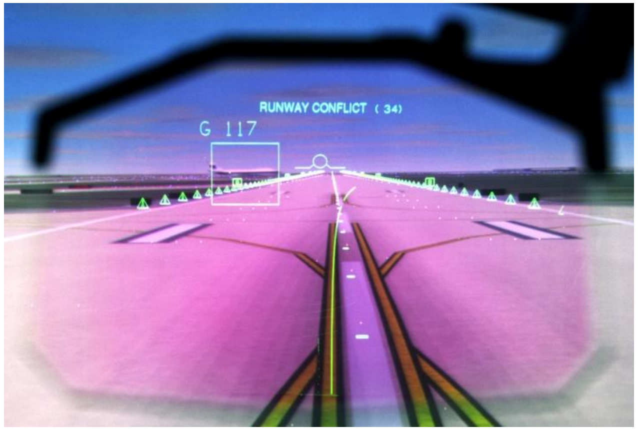
- **The National Transportation Safety Board has runway incursion prevention on its most wanted list for aviation safety**
- **Increase in air traffic forecasted under NextGen could exacerbate this problem**
- **The objective of SESO research is to develop technologies, data, and guidelines to enable conflict detection and resolution in the Terminal Maneuvering Area under NextGen operating concepts providing an additional, protective safety layer**

Traffic position awareness



Departure Surface Map

Ownship position awareness



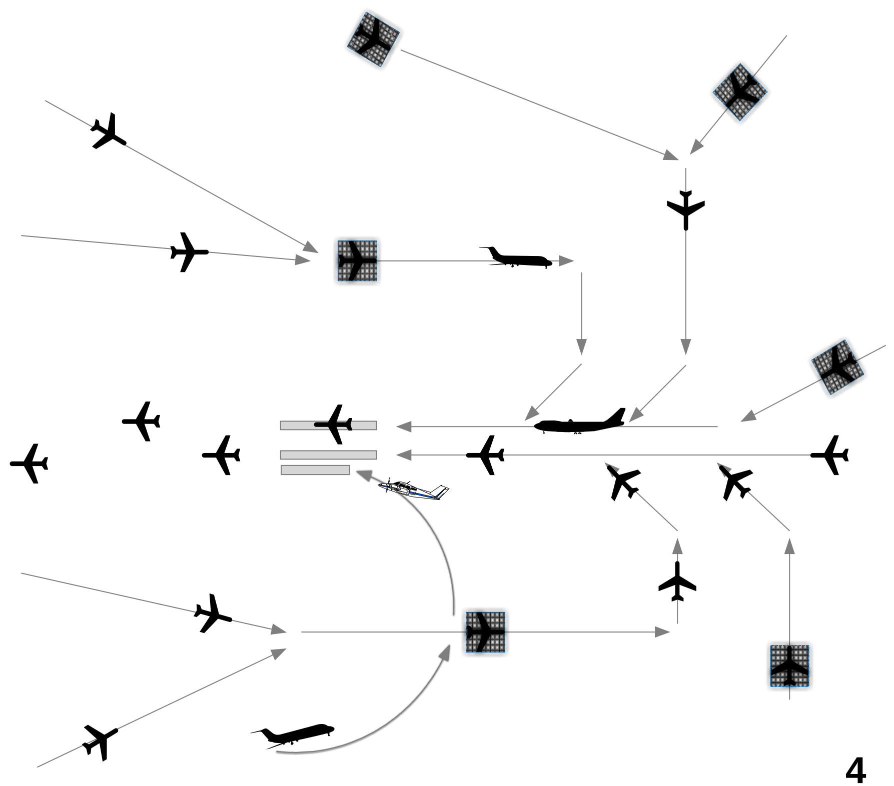
HUD Guidance

Route awareness

Taxi Surface Map

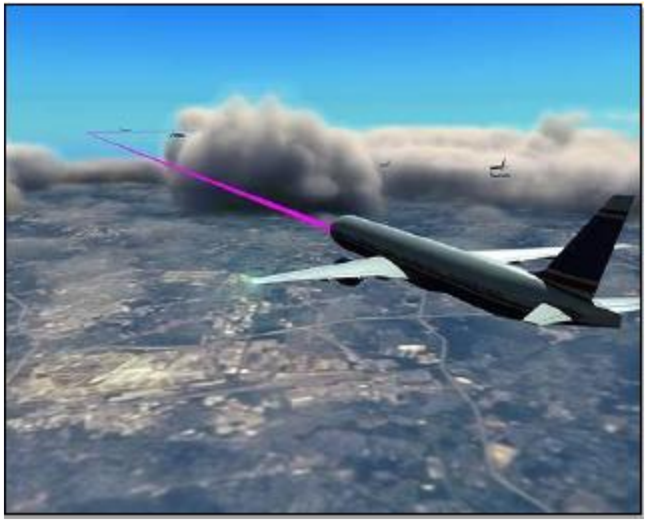
Super Density Operations

- A key to airport efficiency is the ability to schedule, and then manage, the aircraft-to-aircraft spacing at the runway threshold.
- Interleaving complex, three-dimensional routes with time constraints arriving from all directions very difficult for the human mind
- Current scheduling and arrival operations can be made more efficient
- The objective of Super Density Operations (SDO) research is to support the increase in capacity and throughput necessary for NextGen via simultaneous multi-objective sequencing, spacing, merging, and de-confliction for terminal airspace with nearby runway thresholds



Separation Assurance

- The objective of Separation Assurance (SA) research is to develop trajectory-based technologies and human/automation operating concepts capable of safely supporting the increase in capacity necessary for Next Generation Air Transportation System (NextGen)
- In the current Air Traffic Management system, separation of aircraft is the most important task for an air traffic controller in high density airspace and is one of the main factors in controller workload
- This approach is inherently limited by controller workload and will not be able to support the expected traffic growth
- A new airborne trajectory management with self-separation concept developed, in which the pilot is responsible for managing the separation for his or her aircraft supported by onboard automation
- A Human-in-the-Loop (HITL) experiment was conducted to study a new airborne trajectory management with self-separation concept



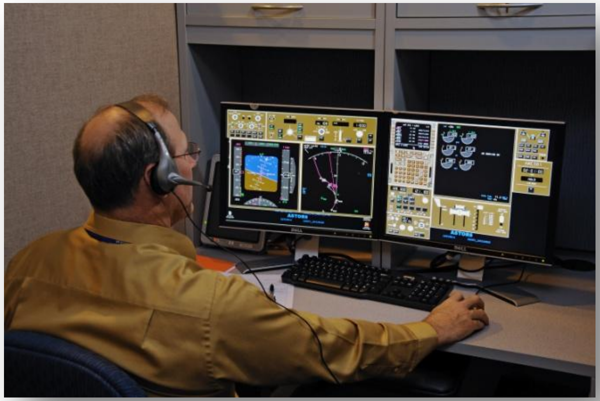
Air Traffic Operations Laboratory

- The Air Traffic Operations Laboratory (ATOL) hosts a simulation platform which provides a medium fidelity setting for studying the interactions of aircraft

- The simulation networks multiple individual pilot stations called Aircraft Simulation for Traffic Operations Research (ASTOR) and a background traffic generator called Traffic Manager (TMX)

- The ATOL has over 300 computers, including 12 desktop pilot workstations, for conducting both Human-in-the-Loop (HITL) and batch (simulation) experiments

Air Traffic Operations Laboratory (ATOL)



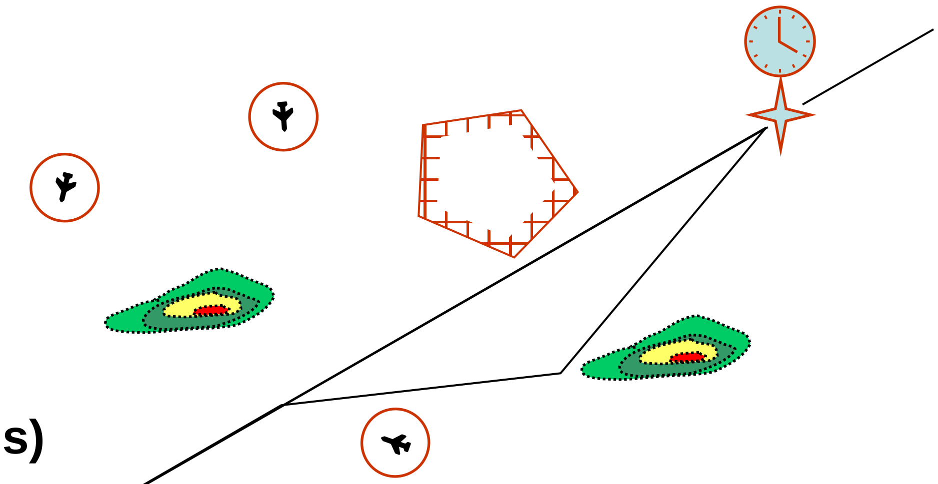
Human Piloted Aircraft

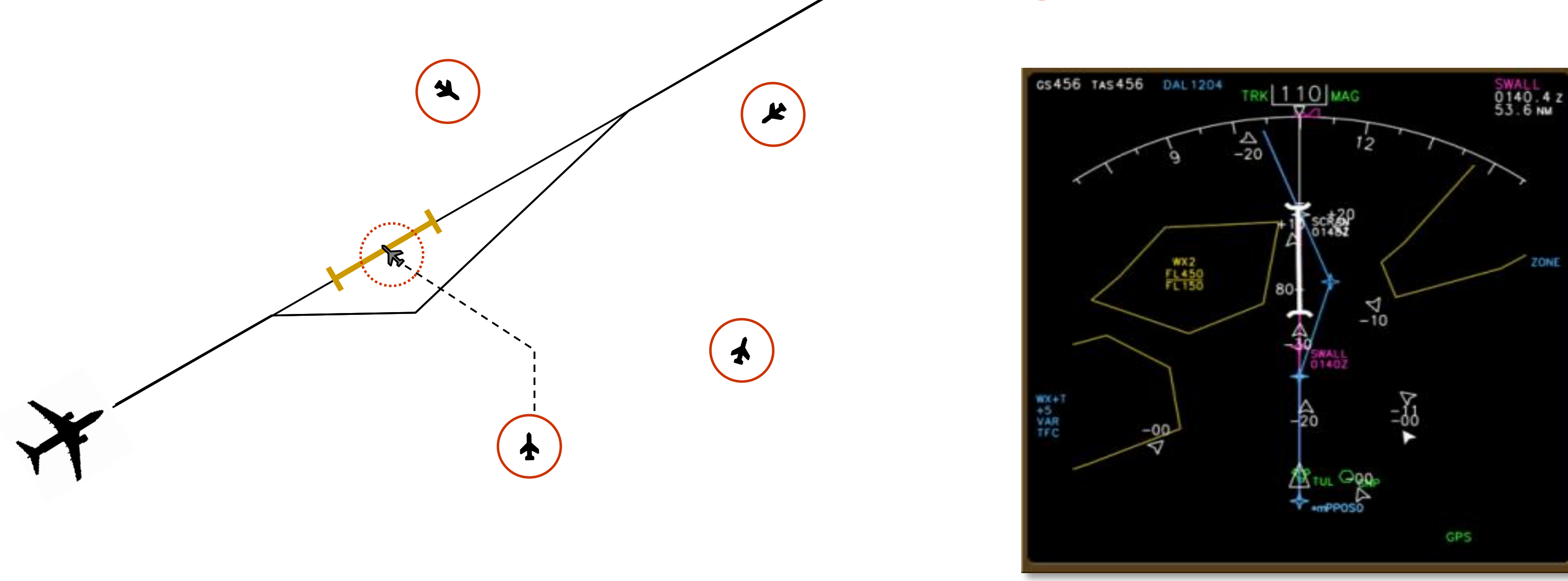
Batch Aircraft (ASTOR & TMX)

Aircraft Simulation for Traffic Operations Research (ASTOR) HITL Interface

Separation Assurance

- Detecting and resolving traffic conflicts
- Verifying all trajectory changes are conflict-free
- Conforming trajectory to additional operational constraints
  - Service provider (e.g. flow constraints, SUA)
  - Operator (e.g. company policy, optimization criteria)
  - Environment (e.g. weather hazards)





Statistical Design of SA HITL

- Research questions are framed as statistical hypotheses to test in the experiment
  - Flight path deviation will be larger in far-term (2.0x) traffic density conditions than in mid-term (1.5x) traffic density conditions

- The experiment is designed to investigate these hypotheses

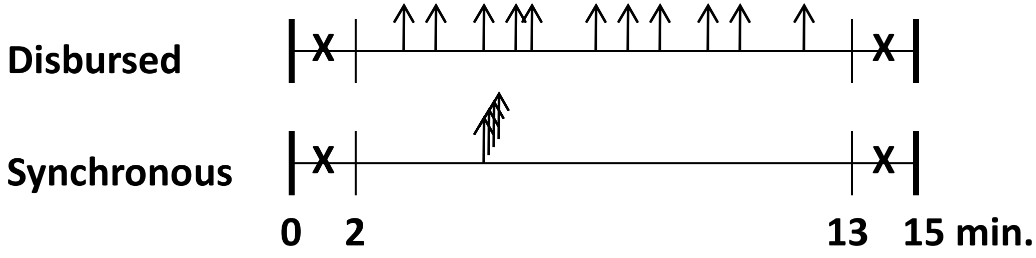
- A formal peer-driven review process is employed to ensure the objectives are met
  - Preliminary Experiment Review (PER)
  - Simulation Requirements Review (SRR)
  - Final Experiment Review (FER)

Experiment Factors / Independent Variables

- Traffic Density (1.5x, 2.0x)
  - Reference: 1x = 18 aircraft per 10,000 nm^2
  - Maintained constant throughout data run

- Scheduling Assignment (No RTA, Yes RTA)
  - Required Time of Arrival (RTA)

- Trajectory Change Event Timing (None, Disbursed, Synchronous)
  - Revised RTA sent via data link
  - Approx 6-8 minute delay requiring path stretch



Experiment Design Matrices

- **30-minute scenarios**
  - **Within-subject design**
  - **No trajectory change events**
  - **2 replicates (8 runs total)**

Scheduling Assignment × **Traffic Density**

Scheduling Assignment	1.5x	2.0x
No	M1	M2
Yes	M4	M3

- **15-minute scenarios**
  - **Within-subject design**
  - **Traffic density (2.0x)**
  - **Scheduling assignment (Yes)**
  - **2 replicates (6 runs total)**

Timing of Trajectory Change Event

	None	*Disbursed*	*Synchronous*
2.0x Traffic Density	S1	S2	S3

Pilot Participants and Experimental Protocol

- 48 pilots: 4 groups of 12 pilots each
  – Groups 1-3: all domestic U.S. pilots
  – Group 4: mix of domestic U.S. and international pilots
    - To support global research perspective on airborne self-separation

- 3 day experiment sessions for each group of participants
  – Day 1: Classroom and hands-on training (10 training scenarios)
  – Day 2: Final training scenario + 8 experiment scenarios (30 min)
  – Day 3: 6 experiment scenarios (15 min) + group debrief session

Blocking Strategy

- Blocking is a method of partitioning the runs into homogeneous sets, or blocks, based on a blocking factor
- Analysis of a block design involves the comparison of runs within the same block, which removes variability due to the blocking factor
- This reduces experimental error and provides more precise answers to research questions
- Group was a blocking factor
  - Four independent groups of 12 pilots participated in four separate three-day experiment sessions
  - Groups 1, 2, and 3 consisted solely of American pilots
  - Group 4 included European pilots to support a global perspective on Air Traffic Management research
  - The groups of pilots were trained separately

Experiment Run Order

- **Order in which treatments are assigned can have important effects on the experiment results and answers to research questions**

- **This is particularly true in HITL experiments where order can affect the behavior of participants due to fatigue, learning curve, or other outside factors**

- **Two of the ways to control for order effects are *randomization* and *counterbalancing***

- **Randomization of the treatments is the most common approach**
  - **Minimizes the impact of any systematic bias on the results**
  - **Is an underlying assumption of most commonly used statistical methods**

Experiment Run Order

- **Counterbalancing assumes a confounding order effect exists which cannot be controlled or randomized out of existence**
  - **Distributes equal amount of the confounding effect to each treatment in such a way that the effect will counterbalance itself and not bias the results**
  - **Main disadvantage is the additional complexity introduced into both the experiment design and data analysis**

- **In the SA HITL, aircraft callsigns were randomly assigned to pilots**
  - **Randomized separately for each run and for each group of pilots**
  - **Scenario difficulty and conflicts encountered varied by callsign**

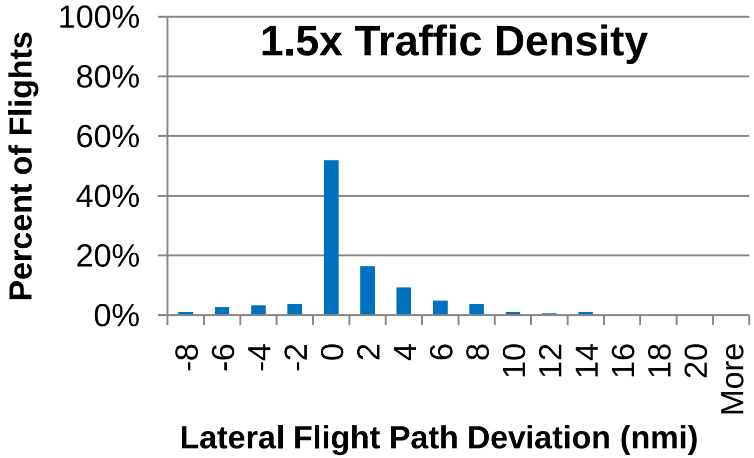
- **Blocked by group of pilots, so the order of the scenarios was randomized separately for each group**

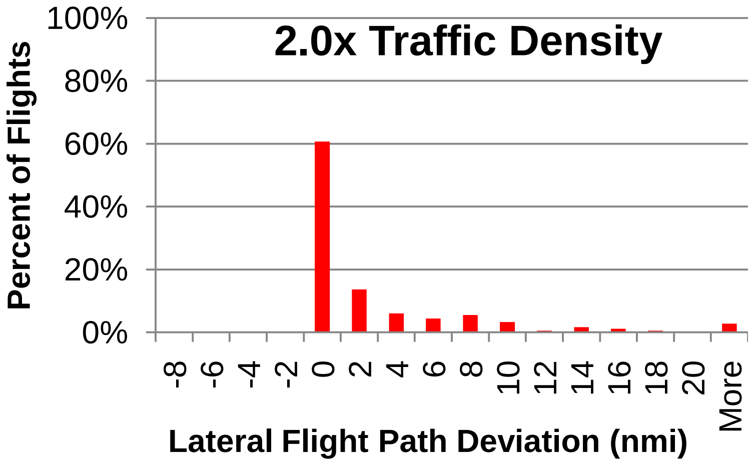
Correlation

- Standard statistical analysis methods are based on the assumption that observations are independent
- However, in HITLs the data have a specific correlation structure
  - Aircraft flown by the same pilot are not independent
  - Aircraft flown by pilots in the same group are not independent
  - Aircraft flown by pilots in different groups are independent
- Three methods for addressing this correlation structure are to ignore it, estimate it, or account for it in the design
  - Ignoring correlation violates the assumption of independence and can lead to over- or under-estimation, which affects all hypothesis tests and conclusions
  - Obtaining a good estimate of the covariance matrix to address the correlation in the data analysis can be very difficult
  - Accounting for the correlation structure in the experiment design is often the best choice, but requires careful planning prior to data collection

Statistical Data Analysis and Interpretation

- **Hypothesis: Flight path deviation will be larger in far-term (2.0x) traffic density conditions than in mid-term (1.5x) traffic density conditions.**
- **Conclusion: The traffic density effect was found to be significant at the alpha = 0.05 level, indicating that increasing the traffic density from 1.5x to 2.0x increased the mean lateral flight path deviation.**





Coordinated Experiments

- **This SA HITL simulation was part of a coordinated experiment with NASA Ames Research Center**
- **The primary goal of these coordinated experiments was to assess the degree of comparability possible**
- **This was the first in a series of experiments within a multi-year research plan to study advanced function allocation concepts for NextGen separation assurance in high density airspace**
- **Although these experiments were jointly designed and conducted in parallel, they differed in the number of replicates, the blocking strategy, and the method for controlling for order effects**
- **Differences in the experiment designs resulted in differences in the data analysis, which made comparison of the results more difficult**

Coordinated Experiments

- **One way to compare the concepts is by conducting statistical hypothesis tests to determine which factors have a significant effect on the response**

- **The quality of a hypothesis test depends on its power, which is the probability of making a correct decision**

- **Initial results can be used to design future coordinated experiments so that statistical hypothesis tests with the same power can be conducted**

- **This would provide a higher degree of comparability for the two concepts**

Conclusions

- **Taking lessons learned from this HITL and other simulations, future experiments will continue to use statistical design of experiments to**
  - **improve efficiency**
  - **answer more research questions with greater precision**
  - **ensure that the experiment design and data collected will allow for the evaluation of the hypotheses of interest**

- **Currently developing experiment plan for the next Separation Assurance HITL investigating a mixed-operation concept**

References

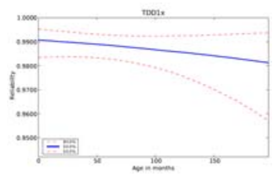
- Wing, D.J., et al., "Function Allocation With Airborne Self-Separation Evaluated in a Piloted Simulation." Proc. 27th ICAS Congress, Nice, France.

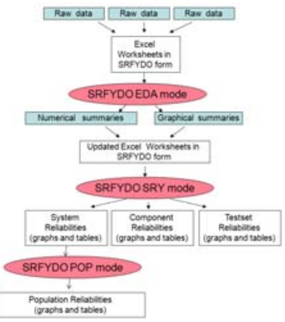
- Consiglio, M.C., et al. (2010). "Human in the Loop Simulation measures of Pilot Response Delay in a Self-Separation Concept of Operations." Proc. 27th ICAS Congress, Nice, France.

- Wing, D.J., et al. (2010). "Comparison of Ground-Based and Airborne Function Allocation Concepts for NextGen Using Human-in-the-Loop Simulations." Proc. 10th AIAA Aviation, Technology, Integration, and Operations (ATIO) Conference, Fort Worth, TX.

LA-UR 11-01594

Modeling the Reliability of Complex Systems with Multiple Data Sources: A Case Study on Making Statistical Tools Accessible to Engineers





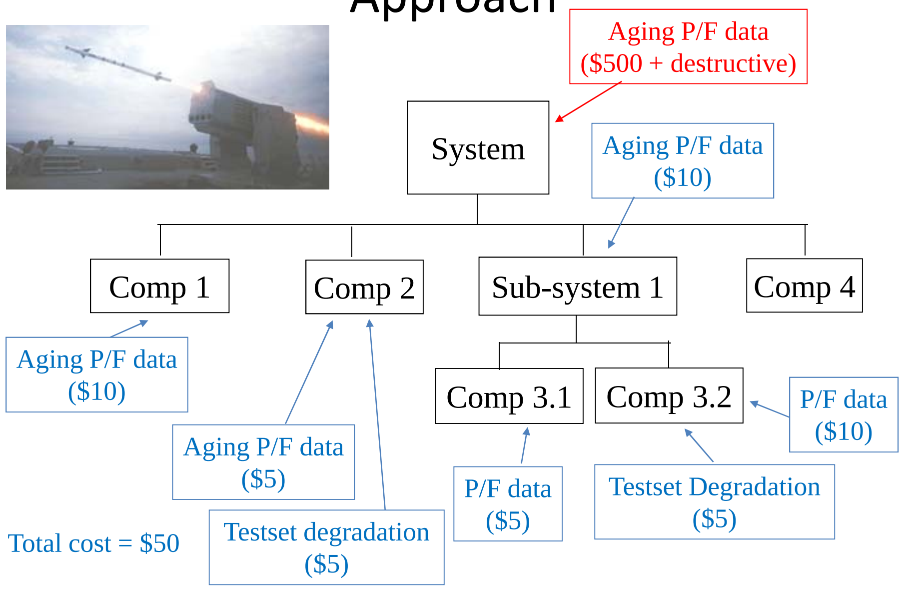
Christine Anderson-Cook
Richard Klamann
Jerome Morzinski
Statistical Sciences Group, Los Alamos National Laboratory

May 2011

Outline

- Motivation for System Reliability Approach – multiple data sources available with expensive full system tests
- New Statistical Method – Bayesian multi-level data combination
- Evolution of SRFYDO (System Reliability Formatter for YADAS Data and Output) Software and Process
- Final Product and Process
- Lessons Learned

Motivation for New System Reliability Approach

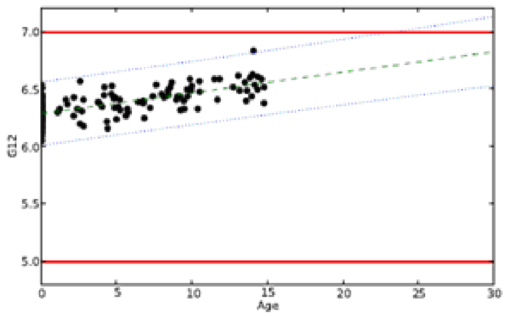


Advantages of SRFYDO Approach

- Uses data already available and thought to be relevant to predict reliability
- Improves precision of estimation with fewer destructive full-system tests
- Check on consistency of information from different data sources
- Flexibility to incorporate partial information into model
- Ability to predict failure before being observed in full-system test
- Component level reliabilities – leverage from different versions of system + better understanding

Disadvantage: More complex statistical method requiring more engineering knowledge to obtain results

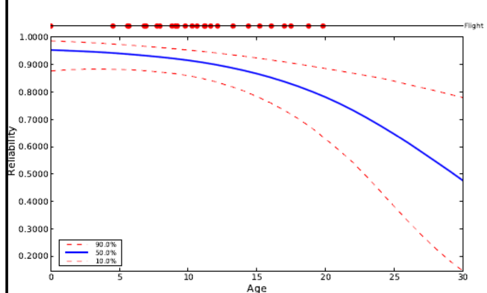
Advantage: Ability to predict failure before being observed in full-system test

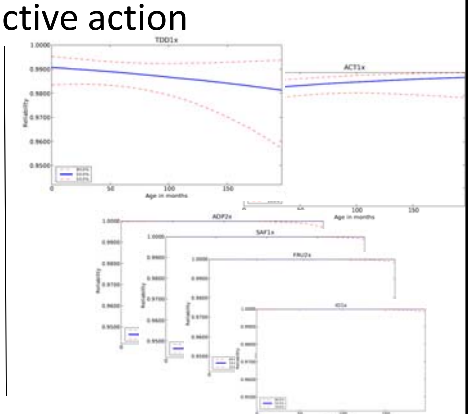


- Because we can track a trend in some of the continuous measurements, we can anticipate when failures might start to occur, before they actually have been observed

Advantage: Component Level Summaries

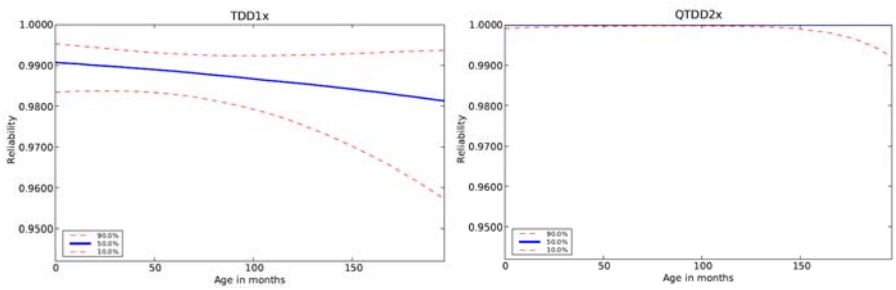
- Better understanding of system and important drivers of system reliability
- Ability to identify critical components and critical specs to implement corrective action





Advantage: Component Level Summaries (cont'd)

- Ability to compare different versions of the same component



Advantage: Component Level Summaries (cont'd)

- Ability to leverage data across different variants with common components
- Data used to estimate reliability:
  - SAF2x + 21 others → 75+47 = 122
  - ADP1x + 6 others → 75
  - ADP2x + 6 others → 47

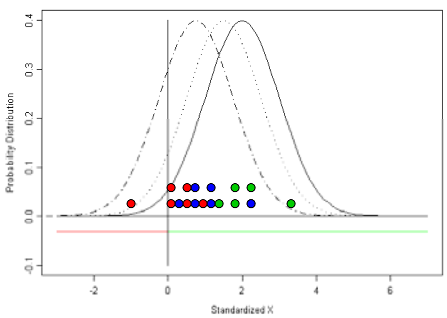
%VariantDef

Version	Component		
Lot4_5	SAF1x	Lot6_7	SAF1x
	WH1x		WH1x
	ACT1x		ACT1x
	BPS1x_Rear		BPS1x_Rear
	RDL1x		RDL1x
	FinSquib1x		FinSquib1x
	MBC1x		MBC1x
	ECU2x		ECU2x
	ADP1x		ADP2x
	Harness1x		Harness1x
	FP1x		FP2x
	FRU1x		FRU2x
75	HH1x	47	HH1x
	IFR1x		IFR2x
	IO1x		IO2x
	LS1x		LS2x
	PM1x		PM1x
	RC1x		RC2x
	RT2x		RT2x
	IRU2x		IRU2x
	TDD1x		TDD1x
	BPS1x_Front		BPS1x_Front
	SS1x		SS1x
	XMTR2x		XMTR2x
	AFD1x		AFD1x
	RocketMotor1x		RocketMotor1x
	RFA1x		RFA1x
	TIVs1x		TIVs1x
	RFP1x		RFP1x

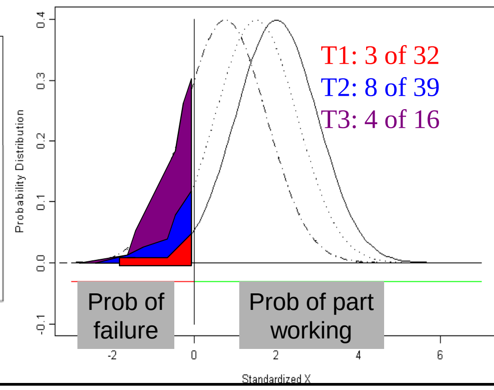
Basic Building Block

- Here we have two potential sources of information about this component:

From testset data, we obtain the mean of the characteristic at each time



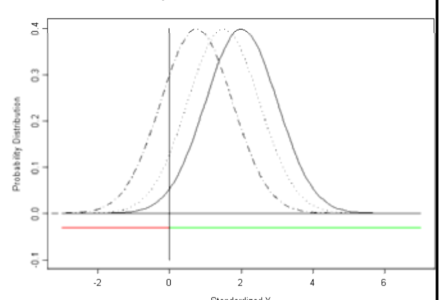
From the full system data, we obtain a proportion of success/failure at each time



Statistical Formulation

- For the probability that a particular component, say component with spec 1, will function correctly

$$p_1(x) = \Phi\left(\frac{\beta_{0,1} + \beta_{1,1}x - \theta_1}{\sqrt{\gamma_1^2 + \sigma_1^2}}\right)$$



Φ = cdf of Normal distribution

$\beta_{0,1}$ – initial mean of testset distribution

$\beta_{1,1}$ – rate of shift of testset distribution

γ_1^2 – variance of testset distribution

θ_1 – discrepancy between means of spec and full system

σ_1^2 – additional variance from full system distribution

Background of Users

- Subject Matter Experts (SME) on particular system
  - System Engineers
  - Data Analysts
- Little or no formal statistical training

- Customers
- Department of Defense
  - NSWC Corona (RAM, ESSM, SeaSPARROW)
  - NSWC Yorktown / Indian Head (AMRAAM)
  - AMCOM/AMRDEC (Hellfire, Stinger)
  - MCPD Fallbrook (TOW)
- Department of Energy
  - LANL Enhanced and Core Surveillance Campaign

Evolution of SRFYDO

1. Development of methods

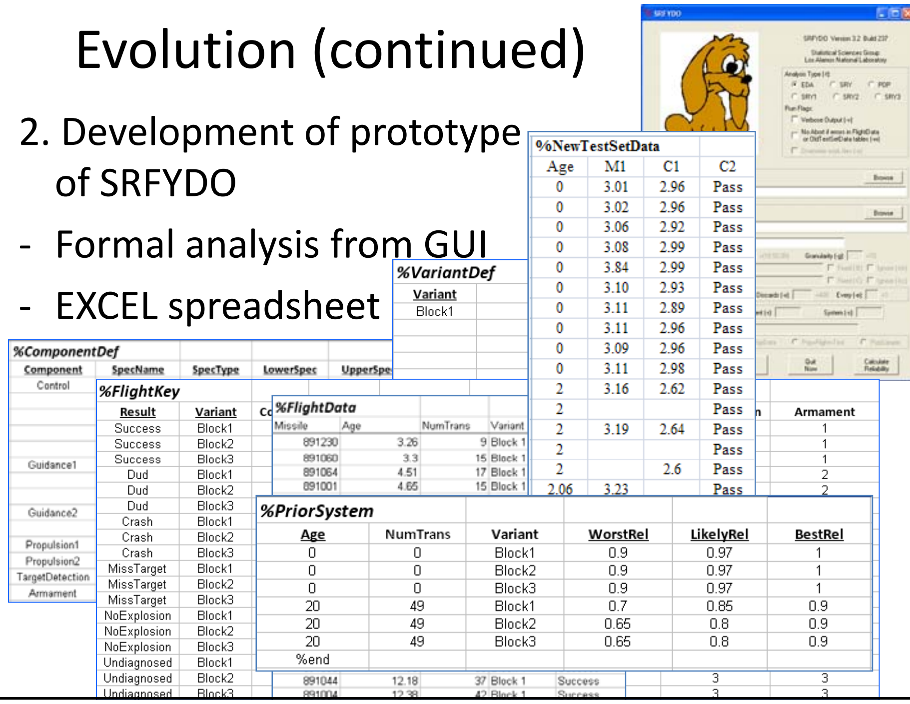
- LANL statisticians sat down with team of SMEs
  - Develop system model (identify components and how connected, map available data to components, obtain priors for analysis)
  - Statisticians did analysis
  - Sat down with SMEs to interpret results, fine-tune model

Characteristics:
- Helpful for development of new methodology – key problems identified
- Long lag for engineers until methods available
- New data added to analysis as it became available
- Methodology implemented with unfriendly code (usable only be creators)
- Very time intensive – not scalable to many systems

Evolution (continued)

2. Development of prototype of SRFYDO

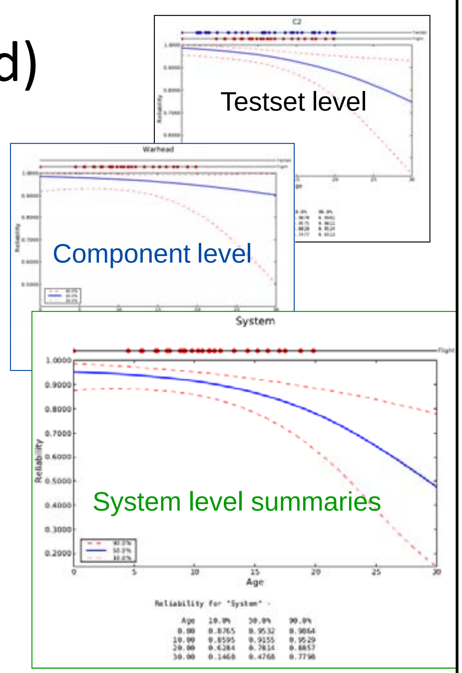
- Formal analysis from GUI
- EXCEL spreadsheet



Age	NumTrans	Variant	WorstRel	LikelyRel	BestRel
0	0	Block1	0.9	0.97	1
0	0	Block2	0.9	0.97	1
0	0	Block3	0.9	0.97	1
20	49	Block1	0.7	0.85	0.9
20	49	Block2	0.65	0.8	0.9
20	49	Block3	0.65	0.8	0.9
%end					

Evolution (cont'd)

- Output as PDF and flat text

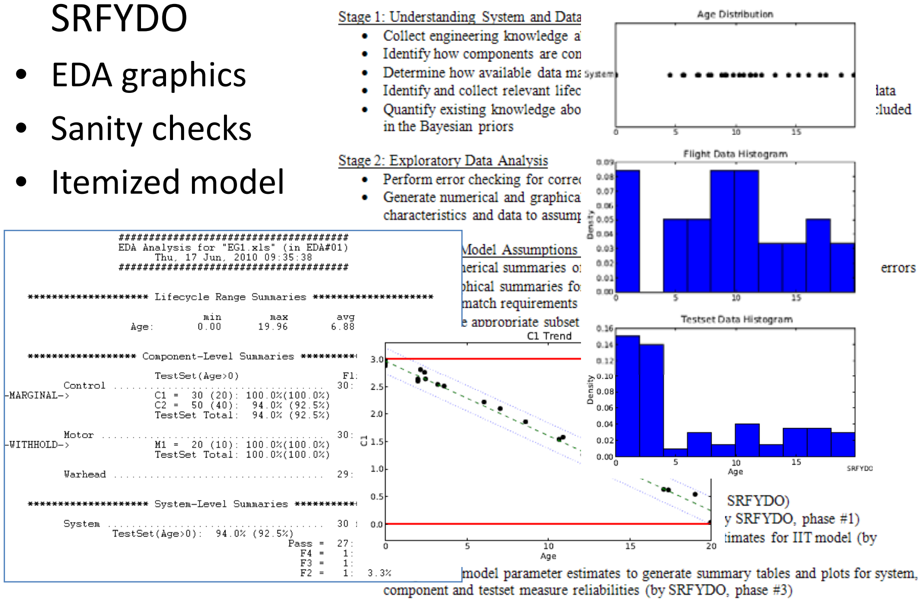


Characteristics:
- SMEs able to function more independently
- Much more timely
- Many requests for special summaries (often integrated into SRFYDO later)

- When applied to new systems, system modeling was often difficult
- Much of data and model assumption checking that LANL did in early stages was not happening (constructing summaries in own software was easy to skip)

3. Larger process developed with EDA stage in SRFYDO

- EDA graphics
- Sanity checks
- Itemized model



The assumptions of the model are listed below:

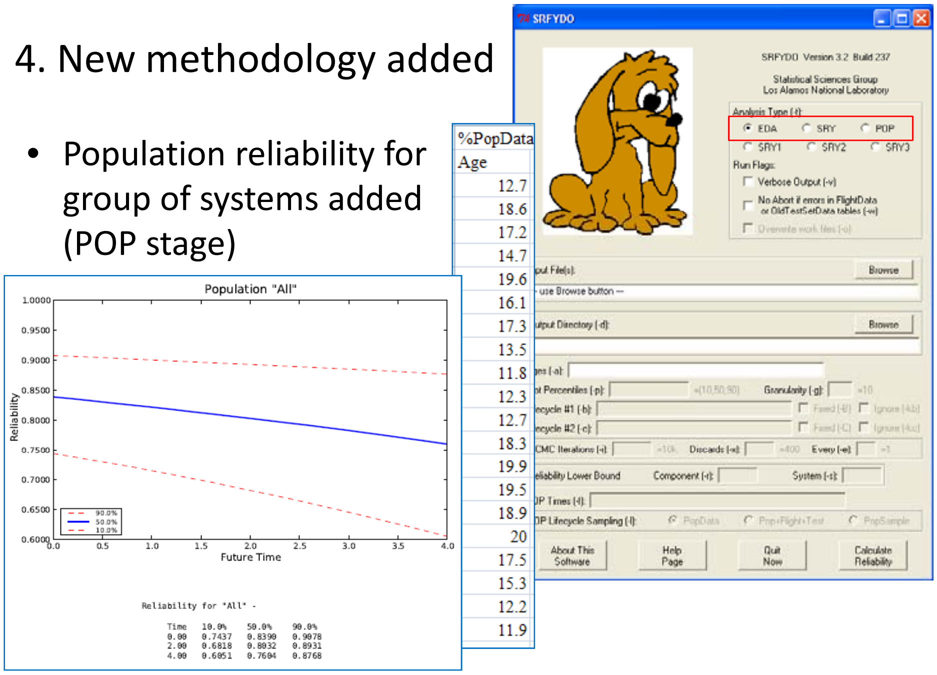
- System Structure
  1. System is a series system.
  2. Only critical testset measures are included in the analysis.
  3. Stockpile of systems is a homogeneous population (or we have lifecycle measures to distinguish between sub-populations.
- Matching Data Types
  4. Full system (flight) tests are considered the most accurate assessment of system reliability.
  5. Surrogacy assumption (systems selected for flight and testset tests have similar lifecycle properties and can be sensibly combined into a single analysis).
  6. Testset data limits correspond to operational limits for what is required of component during a full-system test.
- Testset data:
  7. Linear shift as component ages.
  8. Data at a given time are approximately normally di: symmetric, non-extreme outliers).
  9. Only a single operational limit is important for failu
- Lifecycle covariates
  10. Lifecycle covariates not highly correlated

Process for verifying assumptions:
- Engineering knowledge
- Examining summaries from EDA
- Both

Characteristics:
- SMEs able to function more independently
- Many more discussions about assumptions and boundaries of where model appropriate
- Many fewer re-analyses (huge time-saving)
- More scalable – getting a new system ready for analysis more timely
- SME gaining confidence and expertise with method

4. New methodology added

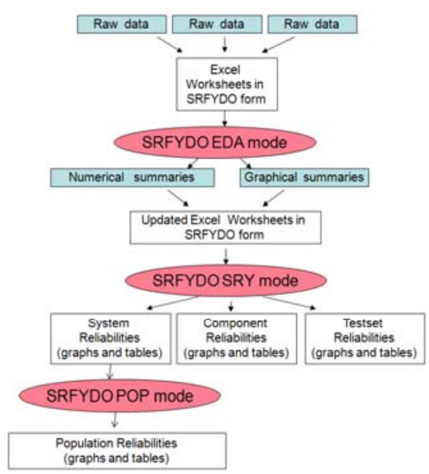
- Population reliability for group of systems added (POP stage)



Final Product and Process

- SRFYDO is the computational engine to guide a process
- EDA mode uses common statistical summaries and graphics which builds in assumptions checking

- Systems analyzed range from:
  - 5 components with one variant
  - 35 components with 8 variants (60+ total components)



Users functioning relatively independently
LANL offers annual training and consulting support

Lessons Learned

- When the focus was on software, our scope was too limited and we were not gaining much traction
- The shift to a guided process (with built in tools for each step) was transformational to our success – the training focuses on the process with SRFYDO being its support
- Assumption check is intuitive for many statisticians, but is built on a foundation of statistical training – making this concrete, accessible and well defined for our customers was essential
- If the summaries / tools needed to perform an analysis are easily available, then the focus shifts to interpretation and decision-making
- The plan evolved and was driven by both the users and the creators

Conclusions

- The process for obtaining system reliability estimates using multiple sources of data using SRFYDO offers a way of incorporating relevant sub-system and component level data to supplement full-system data, which leads to better understanding and a potential improvement to the precision of estimation and prediction
- It allows SMEs to use a sophisticated statistical approach without having to master all of the details of the analysis, but depends of engineering judgment to make sure we have answered the right question

SRFYDO runs on a PC (requires Python, JAVA and Excel) and is available to any US Government agency free of charge srfydo@lanl.gov

Christine Anderson-Cook candcook@lanl.gov

Accelerated Life Testing
Tutorial with NASA and DOD Applications

NASA Statistical Engineering Symposium

Dr. Laura Freeman

Institute for Defense Analyses

May 5, 2011

Outline

- **Introduction**
  - Lifetime data & reliability analysis
  - Weibull distribution
  - Life Tests
  - Accelerated Life Tests (ALTs)
  - Censoring
- **Designing Accelerated Life Tests**
  - Guidelines
  - Monte Carlo Methods
- **Applications**
  - NASA COPV Example
  - Air Force Transponder Mounting Bracket Example

Introduction

- **Reliability: ability of a system to perform a required function**
- **Lifetime data: a quantity of paramount importance to product reliability**
  - Life Tests
  - Accelerated Life Tests
- **Popular distributions for modeling lifetime data**
  - Weibull*
  - Lognormal
  - Exponential
  - Gamma

Weibull Distribution

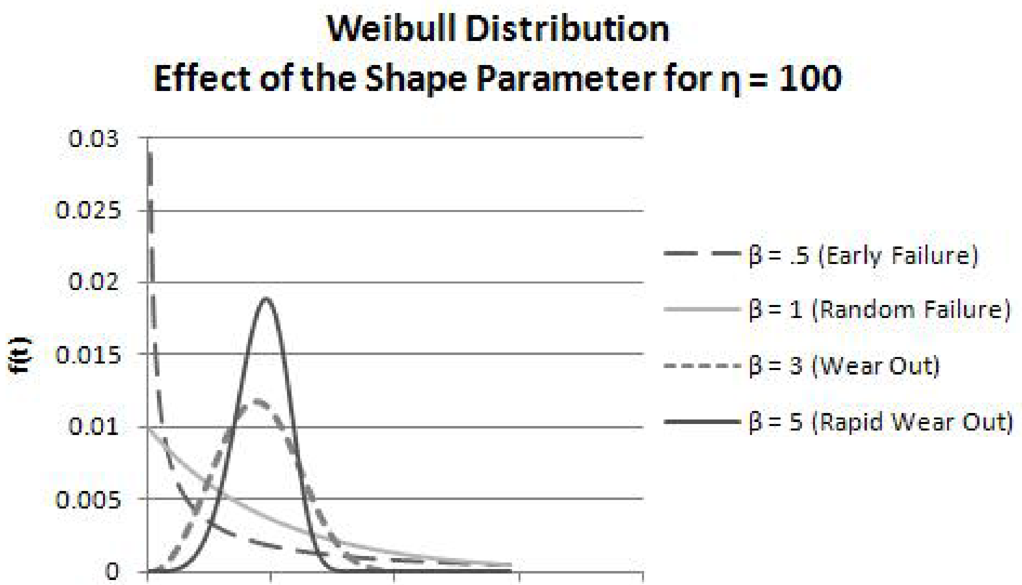
- **Probability density function:**

$$f\left(t,\beta,\eta\right)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

- **Hazard Function:**

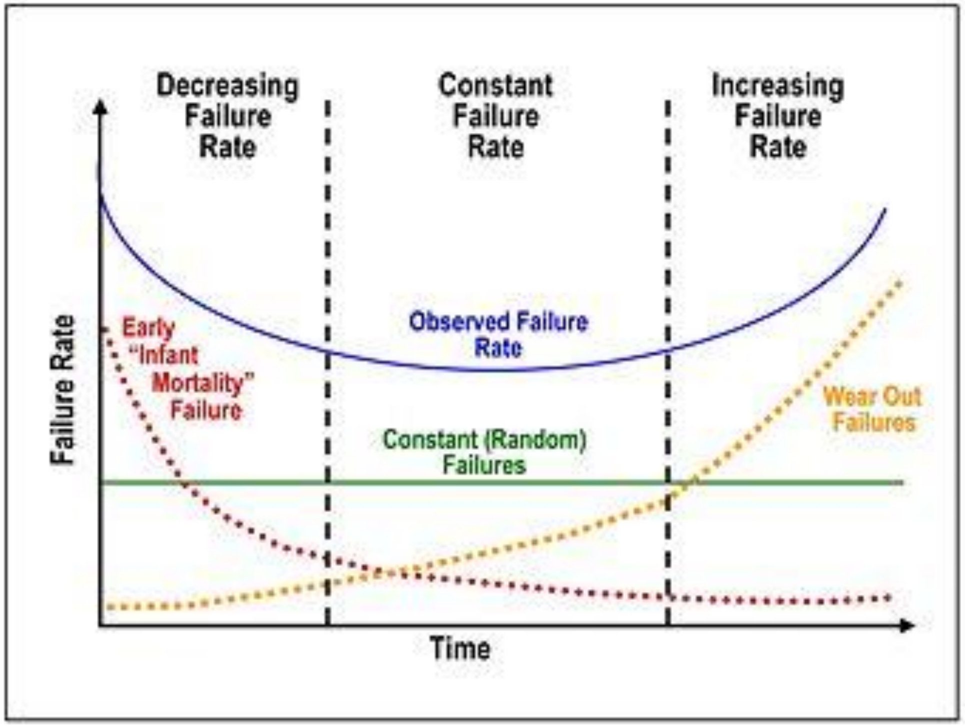
$$h(t)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$$

- **Popular distribution because of its flexibility to model different failure mechanisms**



Weibull Distribution

- **Bathtub Hazard Function**
  - Can be modeled as the mixing of three Weibull distributions.



Designed Experiments & Reliability Testing

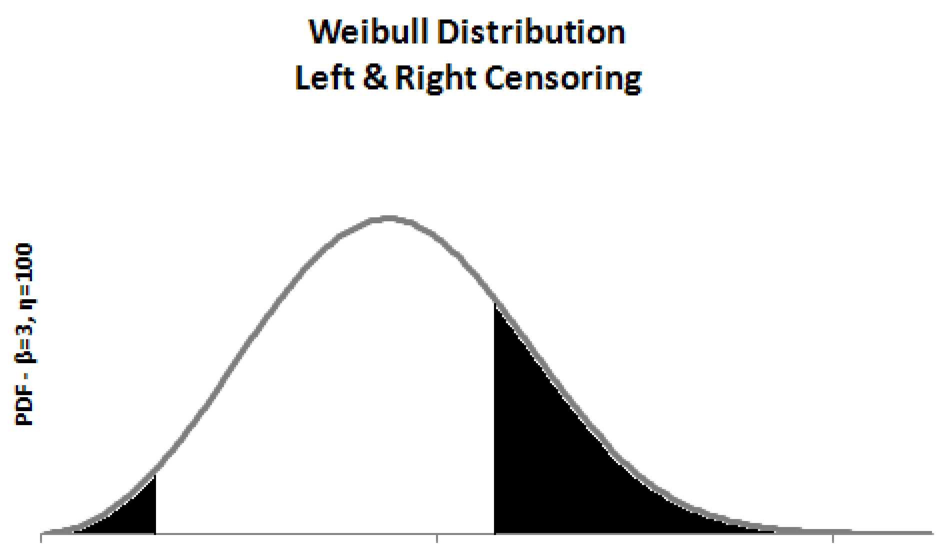
- **Life Tests (LTs)**
  - Goal: model product lifetimes a use conditions
- **Accelerated Life Tests (ALTs)**
  - Goal: Increase the probability of failure by modeling product lifetimes at accelerated conditions
    - » Accelerated in temperature, voltage, humidity, stress, etc.
    - » Project back to use conditions through linearizing relationship
- **Common DOEs for LTs and ALTs**
  - Completely randomized
  - Optimal
  - Designs focus on:
    - » How many units should we use?
    - » How long should we run the test?
- **Complicating issues**
  - Censoring
  - Prediction beyond design space

Censoring

- **Maximum likelihood estimation easily incorporates censoring**

- **Censoring – what is it?**
  - When we are unable to observe a failure time exactly
  - We do know that the unit in question will fail in a certain range

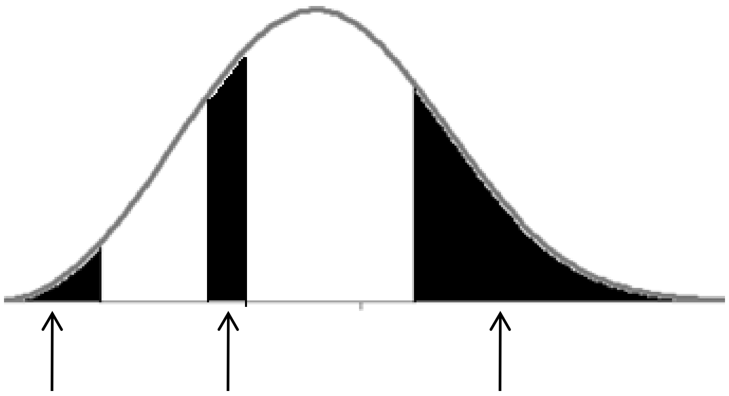
- **Types of Censoring**
  - Left
  - Right (Type I & Type II)
  - Interval



Censoring

- **Contributions to Likelihood**
  - An exact failure time is not observed for a unit
  - Instead we have a range in which the failure occurs
  - Where $F(t_i)$ is the cumulative distribution function at a given time

Censoring Type	Range for Failure Time, T	Likelihood Contribution
Left	$T \leq t_i$	$F(t_i)$
Right	$T \geq t_i$	$1-F(t_i)$
Interval	$t_{i-1} \leq T \leq t_i$	$[F(t_i)-F(t_{i-1})]$
None (Exact Failure)	N/A	$f(t_i)$



Censoring

- **Total Likelihood – product of all likelihood contributions:**

$$L(\theta|y_1,\dots,y_n) = \mathcal{C}\prod_{i=1}^{n} \underbrace{[F(t_i)]^{l_i}}_{\text{Left Censoring Contribution}} \underbrace{[F(t_i)-F(t_{i-1})]^{d_i}}_{\text{Interval Censoring Contribution}} \underbrace{[f(t_i)]^{\delta_i}}_{\text{Exact Failure}} \underbrace{[1-F(t_i)]^{r_i}}_{\text{Right Censoring Contribution}}$$

$$\delta_i = \begin{cases} 1 & \text{if the observation is exact} \\ 0 & \text{if the observation is censored} \end{cases} \qquad l_i = \begin{cases} 1 & \text{if the observation is left censored} \\ 0 & \text{otherwise} \end{cases}$$

$$r_i = \begin{cases} 1 & \text{if the observation is right censored} \\ 0 & \text{otherwise} \end{cases} \qquad d_i = \begin{cases} 1 & \text{if the observation interval censored} \\ 0 & \text{otherwise} \end{cases}$$

Life Tests

- **Designed to measure product lifetime under typical use conditions.**
- **Weibull Model:** $$L(\mu, \sigma; \mathrm{t}) = \mathcal{C} \prod_{i=1}^{n} \left\{ \frac{\beta}{t_i} \phi \left[\beta \left(\log(t_i) - \mu \right) \right] \right\}^{\delta_i} \left\{ 1 - \Phi \left[\beta \left(\log(t_i) - \mu \right) \right] \right\}^{1-\delta_i}$$

$$\delta_i = \begin{cases} 1 & \text{if the observation is exact} \\ 0 & \text{if the observation is censored} \end{cases}$$

$$\log[f(t_i)] = \log\left(\frac{\beta}{t_i}\right) + z_i - \exp(z_i)$$

$$\log[1 - F(t_i)] = -\exp(z_i)$$

$$z_i = \beta \left[\log(t_i) - \mu_i \right] \qquad \mu_i = x_i^T \boldsymbol{\theta} + \epsilon_i$$

- **Limitation**
  – Reliable products may not fail in a reasonable timeframe

Accelerated Life Tests

- **Accelerate the number of failures observed during the test by using one or more accelerating factor**

- **Common methods:**
  - Temperature
  - Stress
  - Humidity

- **Linearizing relationship between model parameters and accelerating variable must be understood.**
  - Engineering knowledge of the relationship is of paramount importance otherwise, model fit will be wrong and projections to use conditions will be nonsensical.

- **Common linearizing relationships:**
  - Arrhenius relationship (temperature)
  - Inverse power law (stress, voltage, pressure acceleration)
  - Generalized Eyring (one or more non-thermal accelerating variables)

Designing Accelerated Life Tests

- **Experimental designs to date focus on:**
  - How many units should we use?
  - How long should we run the test?
  - Under what conditions should I accelerate the units?
- **Prior knowledge of the model parameters is key for planning ALTs**
- **Monte Carlo simulations can be used to construct optimum designs**
  - Minimizing standard error
  - Minimizing the determinant of the Fisher Information matrix
- **Meeker & Escobar recommendations**
  - Caution about using optimum designs without augmentation
  - Use insurance units at use conditions
  - Use 3-4 levels of the accelerating variable
  - Minimize extrapolation (use the lowest level of acceleration possible)
  - Allocate more units to lower levels of the accelerating variable and fewer units to higher levels of the accelerating variable

Applications of Accelerated Life Testing

- **NASA Carbon Fiber Strands for encasing the Composite Overwrapped Pressure Vessel (COPV)**
- **Air Force Transponder Mounting Bracket**

Composite Overwrapped Pressure Vessel (COPV)

- **Problem Statement: Bursting carbon fiber strands is a failure mode that has been observed in the lab but never under use conditions. We need to understand this failure mechanism.**

- **Goal: to develop a model that predicts time to failure for carbon fiber strands at use conditions.**

- **Historical Data:**
  - Kevlar Fiber Strand Testing

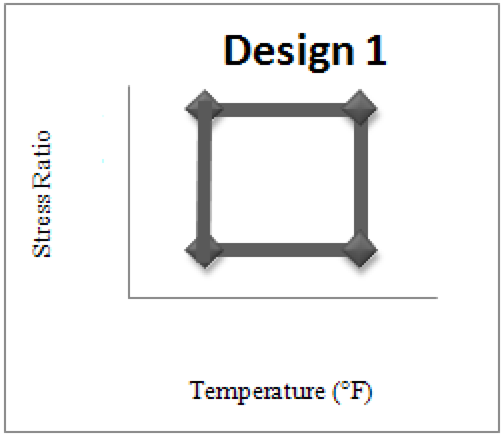
- **Test Approach**
  - Previous data for Kevlar strands focuses on stress ratio acceleration
  - Add temperature acceleration
  - Modified Factorial Design to accommodate ALT specific concerns.

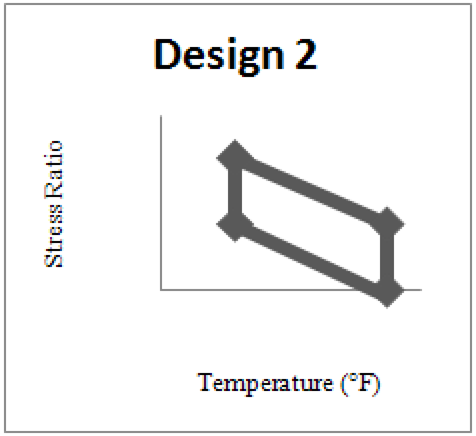
Composite Overwrapped Pressure Vessel (COPV)

- **Classic Power Law model:** $$F(t_i) = 1 - \exp\left[-\left[\frac{t_i}{t_{ref}} SR^{\rho}\right]^{\beta}\right]$$

$$F(t_i) = 1 - \exp\left[-\left[\frac{t_i}{\exp(\gamma_0)} SR^{-\gamma_1}\right]^{\beta}\right]$$

- **Weibull Model:** $$F(t_i) = 1 - \exp\left[-\left[\frac{t_i}{\exp(\gamma_0 + \gamma_1 \log(SR))}\right]^{\beta}\right]$$

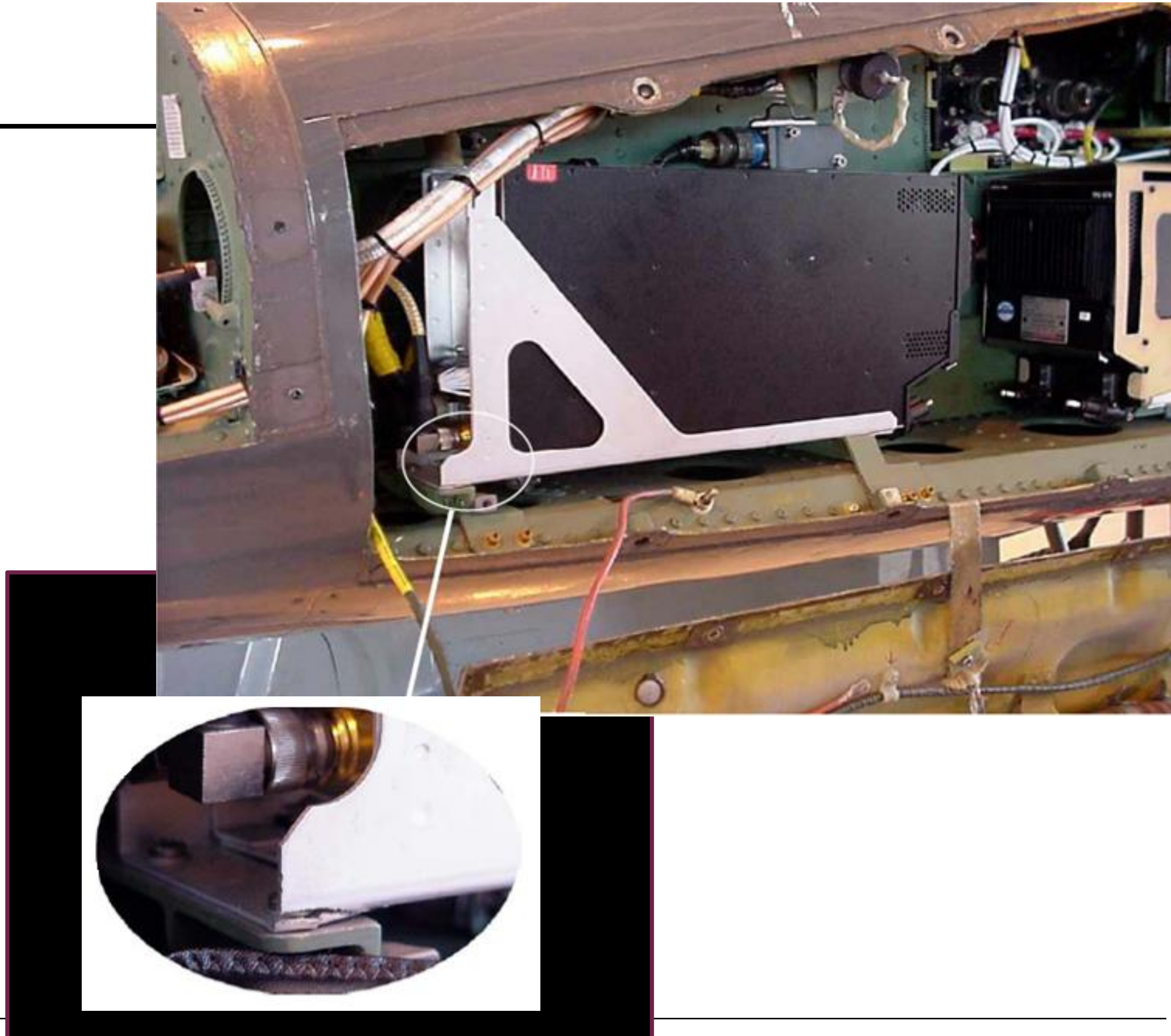




Stress Ratio (SR)	Temp (°F)	Number of Strands	Expected Number of Failures at One Week
Low	High	25	4.49
Medium	Low	25	11.72
Medium	High	15	5.11
High	Low	15	9.25
Total Number		80	

Mounting Bracket for Aircraft Transponder Tray

- **Problem Statement:**
  - The mounting bracket that holds the transponder tray in place on military aircraft are cracking. They were designed to be used on commercial aircraft. To fix the problem the Air Force has proposed an updated mounting tray with an extra stabilizer. However, there is concern that this additional stabilizer may induce a new failure mechanism.
- **Goal: to develop a model that predicts time to failure for the new mounting bracket.**
- **Historical Data:**
  - Time to failure for historical mounting bracket.
  - Times are interval censored.
- **Test Approach**
  - Vibration Acceleration
  - For operational realism, mounting bracket needs to be tested with actual aircraft and transponder tray.



Applications & Challenges in DoD Testing

- **Need for ALT Application in DoD Testing**
  - Nearly all military systems have reliability requirements that are not achievable in the typical test period.
  - Increased emphasis on reliability.
  - Upgrades to existing systems.

- **Challenges**
  - General caution about statistical models, they have not been differentiated from modeling and simulation.
  - Projection beyond the test design space caries increased risk.
  - Limited capabilities to implement these types of statistical methodologies in DoD testing.

References

- **Textbooks:**
  - Lawless, J. F. (2002). *Statistical Models and Methods for Lifetime Data.* Hoboken, New Jersey: John Wiley & Sons Inc.
  - Meeker, W. Q., & Escobar, L. (1998). *Statistical Methods for Reliability Data.* New York: John Wiley & Sons Inc.
  - Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analysis.* New York: John Wiley and Sons.

- **Kevlar Fiber Strand Papers:**
  - Feiveson, A., and Kulkarni, P. Reliability of Space-Shuttle Pressure Vessels with Random Batch Effect. Technometrics 42, 4 (2000), 332{344.
  - Leon, R., Ramachandran, R., Ashby, A., and Thyagarajan, J. Bayesian Modeling of Accelerated Life Tests with Random Effect. Journal of Quality Technology39, 1 (2009), 3-16.

Responding to Climate Variability and Change: A Rapid Prototype For Assessing Impacts of Uncertainty in Climate Observations and Model Projections on Decision Support

May, 2011

Lead:
L. DeWayne Cecil, Chief for Science Applications, USGS Global Change Program Office
Research Team:
NASA GISS: Cynthia Rosenzweig, Radley Horton, Alex Ruane
NASA Langley: Peter Parker and Brian Killough
Booz Allen Hamilton: Ray McCollum, Douglas Brown
Stakeholders:
Regional Committee on Hydraulic Resources
Regional Monitoring and Visualization System for Mesoamerica

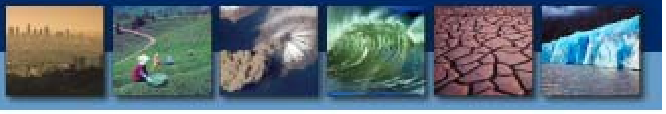
NASA | USGS science for a changing world | Booz | Allen | Hamilton

CEOS Committee on Earth Observation Satellites

Outline

- ❖Sources Of Uncertainty in Climate Science and Applications
- ❖Crop Yield Prediction Process
- ❖ Analysis Of Results
- ❖ Uncertainty Characterization and Communication

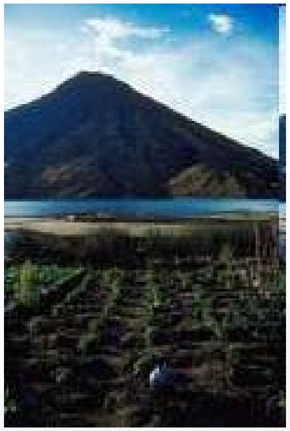
Committee On Earth Observation Satellites (CEOS) Purpose & Scope



Societal Benefit

Decision Topic

What is the impact of climate change (temperature and precipitation) and its uncertainty on the change to agriculture yield in Central America and how will it impact water resource management decisions in that region?



Information Products and Services

- Temperature Forecasts
- Precipitation Forecasts
- Decision Support Tools

IPCC Scenarios

SERVIR and DSSAT – Central America

Uncertainty Analysis

Examine the uncertainty at every level (measurements to forecasts) and determine the impact on a specific decision.

Science Knowledge and Models

- Regional Climate Models (RCM)
- Global Circulation Models (GCM)

WRF – Central America

- NCAR Community Climate System Model (CCSM)

Precipitation
Temperature

Measurements

Earth observation satellites

Instruments and Missions

NASA | USGS science for a changing world | Booz | Allen | Hamilton

Sources of Uncertainty Analyzed

❖ **Climate Uncertainty**
- **Rainfall uncertainty regarding the effect of rainfall on maize yield**
- **Temperature uncertainty regarding the effect of temperature on maize yield**

❖ **Emission Uncertainty**
- **Two emission scenarios affect the maize yield, rainfall, and temperature predictions for all of the models**

❖ **Model Uncertainty**
- **Global Circulation Model (GCM) Uncertainties**
  - o **Uncertainty exists from model to model in that each model makes distinct predictions about maize yield, rainfall, and temperature**

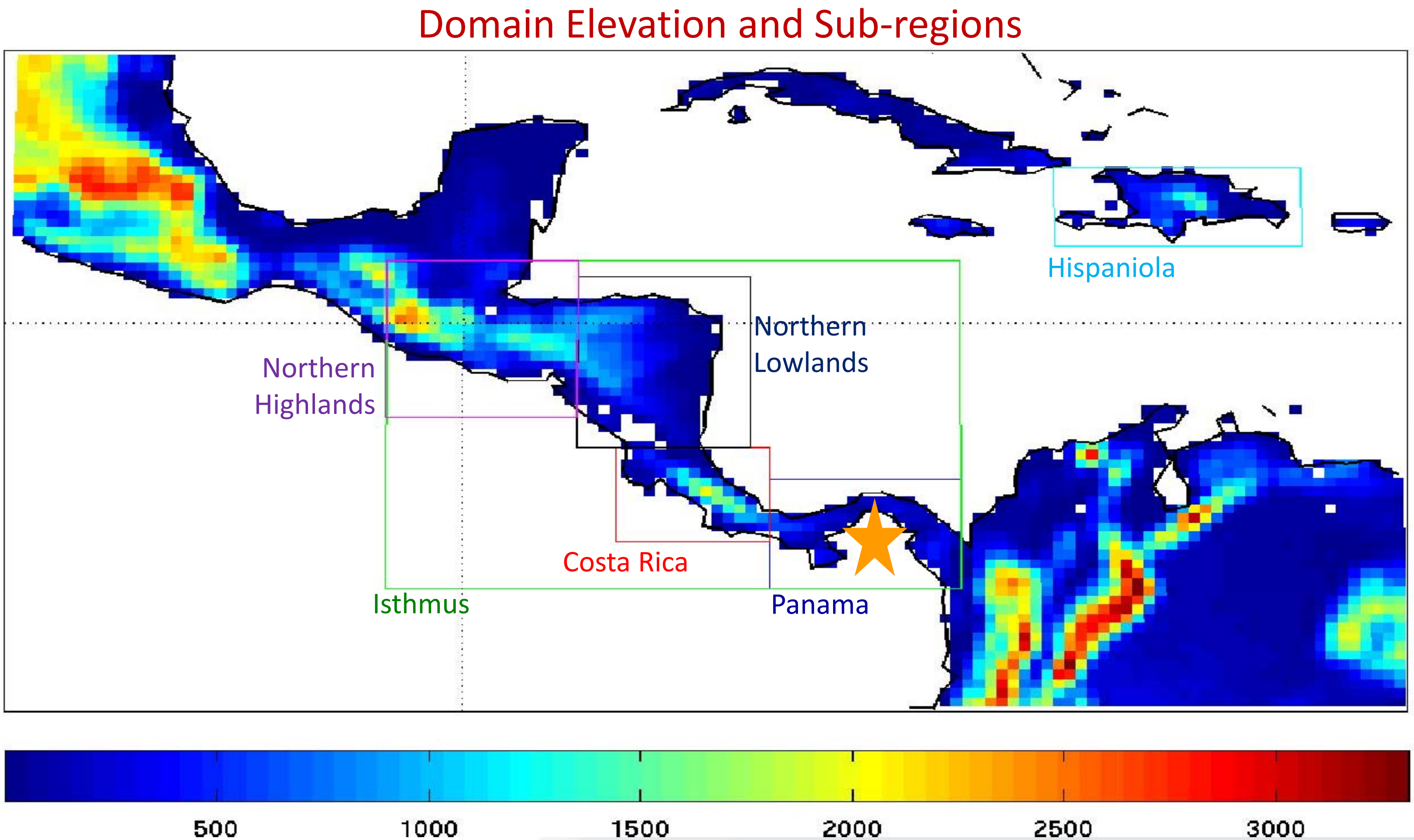
Goals for Communicating and Displaying Uncertainties for Decision Support

- Develop a repeatable and traceable framework for characterizing uncertainty in climate modeling
- Isolate the few important factors impacting decision support from the trivial many
- Focus on incorporating all available information and sources rather than subjectively biased selections
- Incorporate insightful graphical communication that enables exploratory analysis of critical factors
- Provide answers to decision-makers that clearly communicate uncertainty and risk

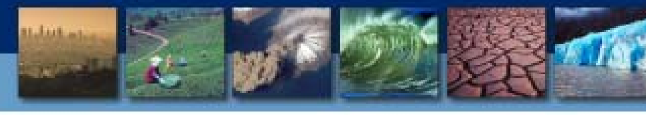
Outline

- ❖ Sources Of Uncertainty in Climate Science and Applications
- ❖ Crop Yield Prediction Process
- ❖ Analysis Of Results
- ❖ Uncertainty Characterization and Communication

Focus Region For The Pilot Study



Crop Yield Prediction Process - Overview



- ❖ The crop yield prediction process requires a synthesis of models and information:
  - Tocumen, Panama (located near Panama City), was selected as a trial location
  - This site was selected because of its longer history of daily meteorological observations
  - We generated crop yield data for maize based upon the characteristics of local fields, their agricultural practices, and observed meteorological conditions
  - Scenarios of climate change during the planting season were generated for the region
  - Yield simulations driven by historic and future climate scenarios are compared to determine the range of likely impacts of climate change.

Crop Yield Prediction Process – Crop Model

- Maize yield was simulated using the Crop Estimation through Resource and Environment Synthesis (CERES-Maize) biophysical crop model, which is an element of the Decision Support System for Agrotechnology Transfer (DSSAT) family of models

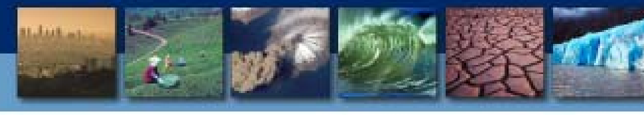
- CERES-Maize requires the following information:
  - Crop cultivar: genetic information about the crop to be grown
  - Field characteristics: soil type, initial moisture, field drainage, etc.
  - Agricultural management: planting dates and geometry, irrigation/fertilizer applications, etc.
  - Local meteorology: daily sunshine, min/max temperatures, and rainfall

Crop Yield Prediction Process – Future Climate

- Climate change information was taken by comparing future (2020-2049) and baseline (1970-1999) climates from five climate models contributed to the Intergovernmental Panel on Climate Change (IPCC) 4th Assessment Report.
  - The A2 and B1 scenarios were examined to gauge societal uncertainties
    - For definitions of A2 and B1 see *IPCC Special Report Emissions Scenarios*
  - Monthly temperature changes
  - Monthly rainfall percentage changes
  - Historic daily meteorology and carbon dioxide concentrations modified according to climate changes to produce future scenarios

GCM Name	Institution	Atmospheric Resolution (lat, lon, °)	Equilibrium Climate Sensitivity (°C) for doubling of CO_2
gfdl_cm2_1	Geophysical Fluid Dynamics Laboratory, USA	2x2.5	3.4
giss_model_e_r	NASA Goddard Institute for Space Studies, USA	4x5	2.7
mpi_echam5	Max Planck Institute for Meteorology, Germany	1.878x1.88	3.4
ncar_ccsm3_0	University Corporation for Atmospheric Research, USA	1.4x1.4	2.7
ukmo_hadcm3	Hadley Centre for Climate Prediction, Met Office, UK	2.5x3.75	3.3

Rainfall & Temperature Projections Are Aggregated To Seasonal Periods Over The Prediction Range (2020-2049)



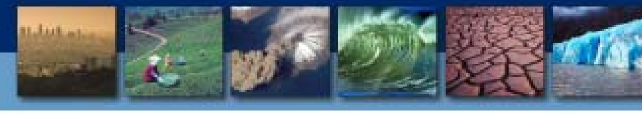
- Temperature is measured in degrees Celsius
  - Chart represents the change in temperature from the baseline period
- Rainfall is measured in millimeters
  - Chart represents the percentage change in rainfall from the baseline period

Temp Change	GFDL2.1_A2	GISS_Er_A2	ECHAM5_A2	CCSM_A2	UKMO_A2	GFDL2.1_B1	GISS_Er_B1	ECHAM5_B1	CCSM_B1	UKMO_B1
May	1.91	1.79	1.12	1.21	1.56	1.27	1.43	1.23	0.83	1.73
June	1.36	1.45	0.96	1.25	1.19	1.07	1.26	1.12	0.94	1.4
July	1.24	1.18	0.88	1.22	1.2	1.1	1.1	0.8	0.98	1.07
August	1.29	1.28	0.57	1.27	0.85	1.22	1.18	0.54	1.06	0.65
September	1.18	1.34	0.53	1.3	1.02	1.06	1.08	0.57	1.02	0.95

Rainfall Change	GFDL2.1_A2	GISS_Er_A2	ECHAM5_A2	CCSM_A2	UKMO_A2	GFDL2.1_B1	GISS_Er_B1	ECHAM5_B1	CCSM_B1	UKMO_B1
May	-4.6	-1.2	0.2	3.2	-5.5	18.6	-0.8	-8	-5.8	-13.5
June	-4.3	-5.5	1	0.8	18.7	-17.3	-5.1	0.2	2.1	7.7
July	-13.5	2.9	7.3	-7	0.6	-14.5	2	4.6	-4.1	2.4
August	-10.1	3	6.4	-4.9	7.5	4.6	-1.7	4.1	-1.7	1.2
September	-10.4	-2.6	0.1	-0.4	13.8	-2	-1	-7.2	6.3	3.3

NASA USGS science for a changing world Booz | Allen | Hamilton

Five GCMs Project Maize Yield Through IPCC Emission Scenarios A2 And B1



baseline yr	baseline	GCM yr	GFDL2.1_A2	GISS_Er_A2	ECHAM5_A2	CCSM_A2	UKMO_A2	GFDL2.1_B1	GISS_Er_B1	ECHAM5_B1	CCSM_B1	UKMO_B1
1970	4264	2020	3192	3397	3758	2926	3556	3122	3331	3664	3433	3589
1971	3026	2021	3418	3537	4321	3848	3910	3401	3621	4088	3810	3627
1972	5296	2022	5310	5537	5768	4980	5279	4932	5109	5410	5224	5588
1973	3922	2023	3273	3436	3774	3397	3547	3261	3378	3603	3553	3477
1974	3205	2024	3778	3714	4151	3639	3772	3803	3684	3994	3595	3868
1975	3813	2025	4267	4437	4784	4284	4572	4165	4369	4223	4146	4490
1976	2873	2026	3506	3747	3214	4014	4120	3846	3987	3044	3113	3805
1977	3086	2027	3624	3714	3396	3028	3785	3126	3672	3436	3174	3816
1978	4781	2028	4689	4863	5127	4550	4772	4569	4618	5223	4963	5025
1979	4150	2029	3813	3998	4504	3969	4306	3827	3880	4439	3864	4357
1980	1785	2030	2911	2982	2509	2198	2408	2306	2316	2249	2540	3106
1981	4213	2031	4142	4099	4341	3781	4409	3706	3933	4186	3782	4423
1982	3757	2032	3539	4031	4528	4069	4095	3648	3994	4238	4100	4104
1983	2635	2033	3170	3371	3762	3114	3364	3224	3366	3616	3062	3478
1984	3615	2034	3929	4092	3953	3514	4185	3468	4077	3755	3637	4054
1985	2490	2035	3119	3439	3410	3105	3314	2972	3147	3113	2975	3186
1986	1746	2036	3301	3087	3347	3322	3255	3029	3041	3316	2957	3149
1987	4012	2037	3922	4005	3704	3807	4521	3751	3738	3820	3801	4106

- **Yield is measured in kilograms per 10,000 square meters**

GCM

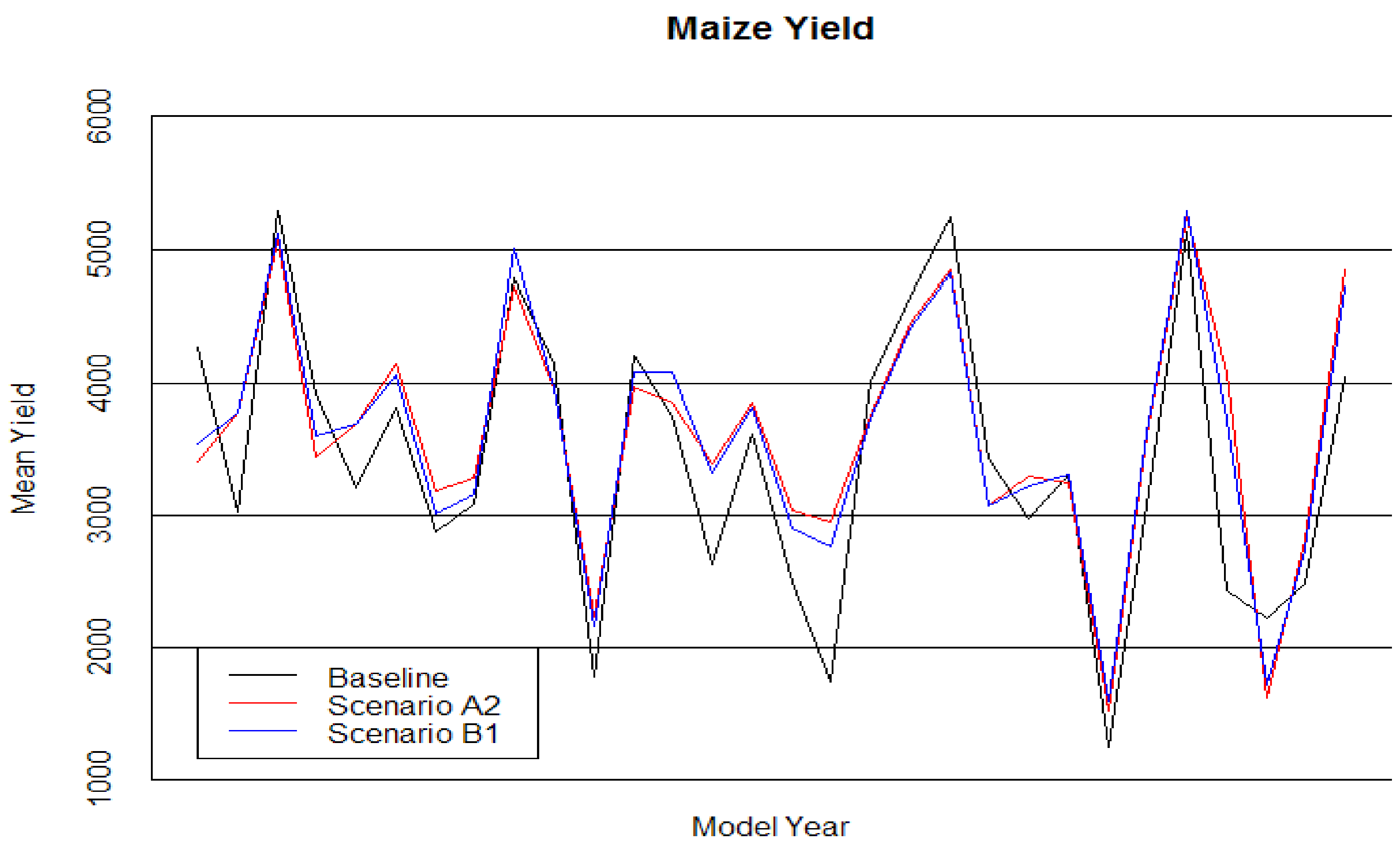
Baseline Observation

Time

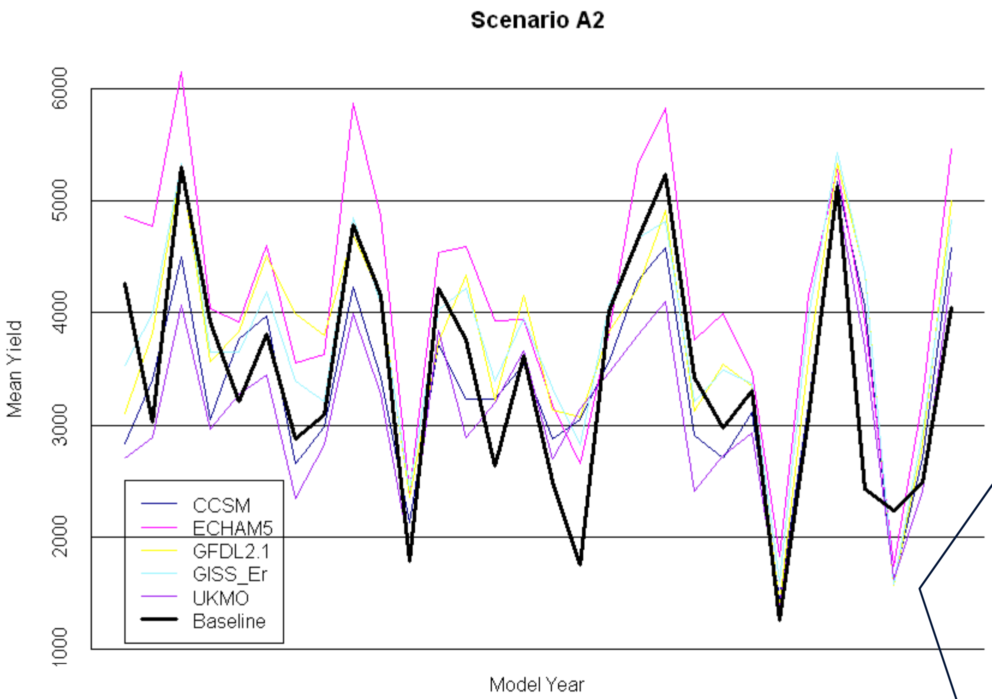
Outline

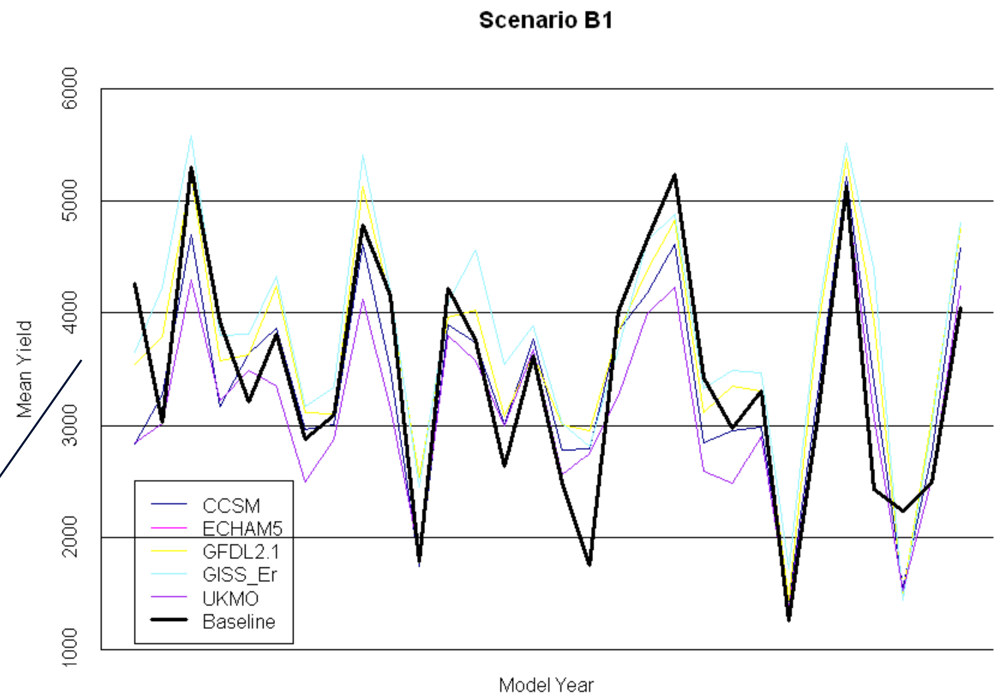
- ❖ Overview of CEOS and the Project
- ❖ Sources Of Uncertainty in Climate Science and Applications
- ❖ Crop Yield Prediction Process
- ❖ Analysis Of Results
- ❖ Uncertainty Characterization and Communication

Baseline Maize Yield & GCM-Projected Maize Yield (2020-2049) By Scenario Across GCMs



GCM-Projected (2020-2049) Maize Yield By Model & Scenario

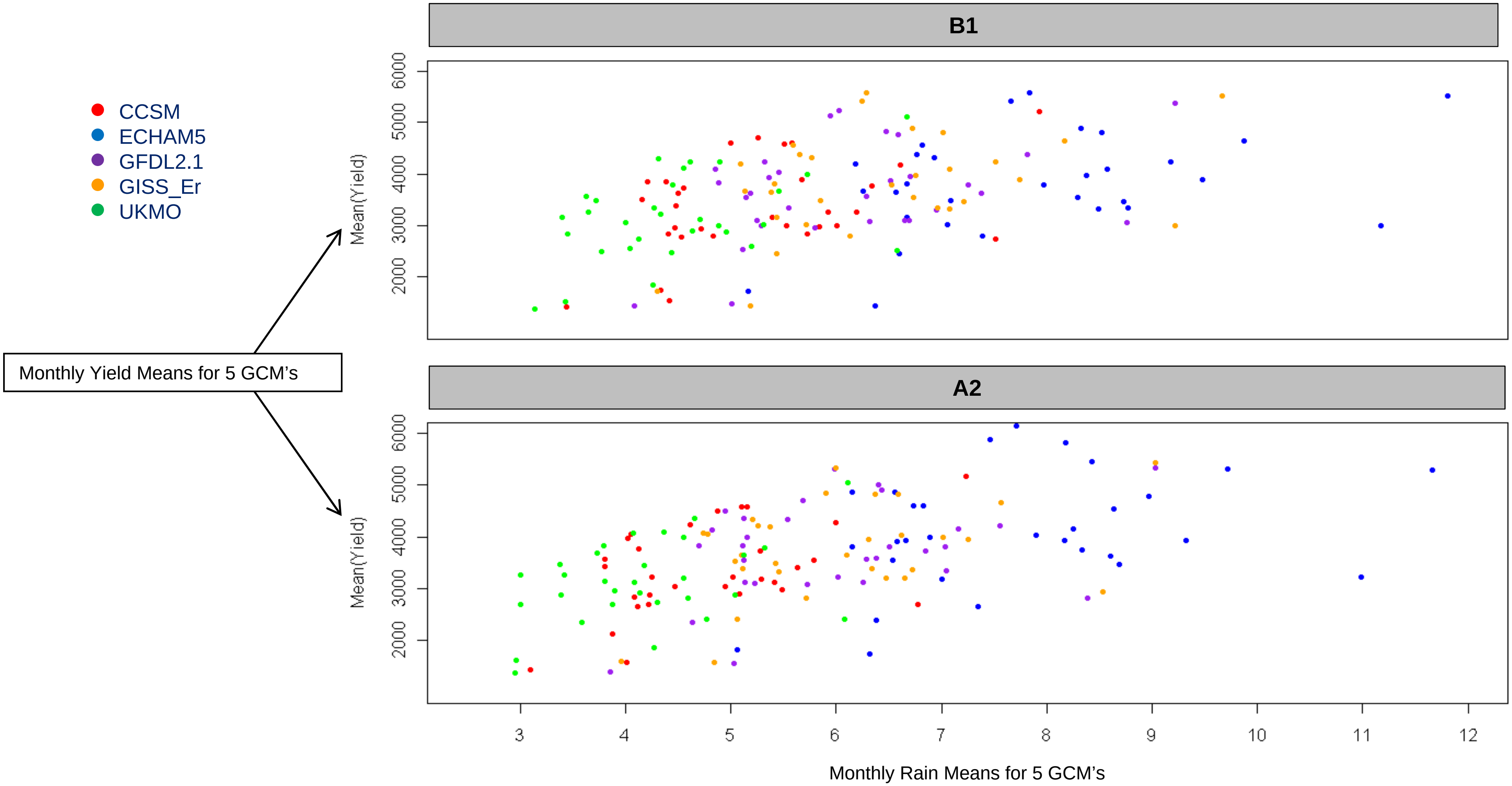




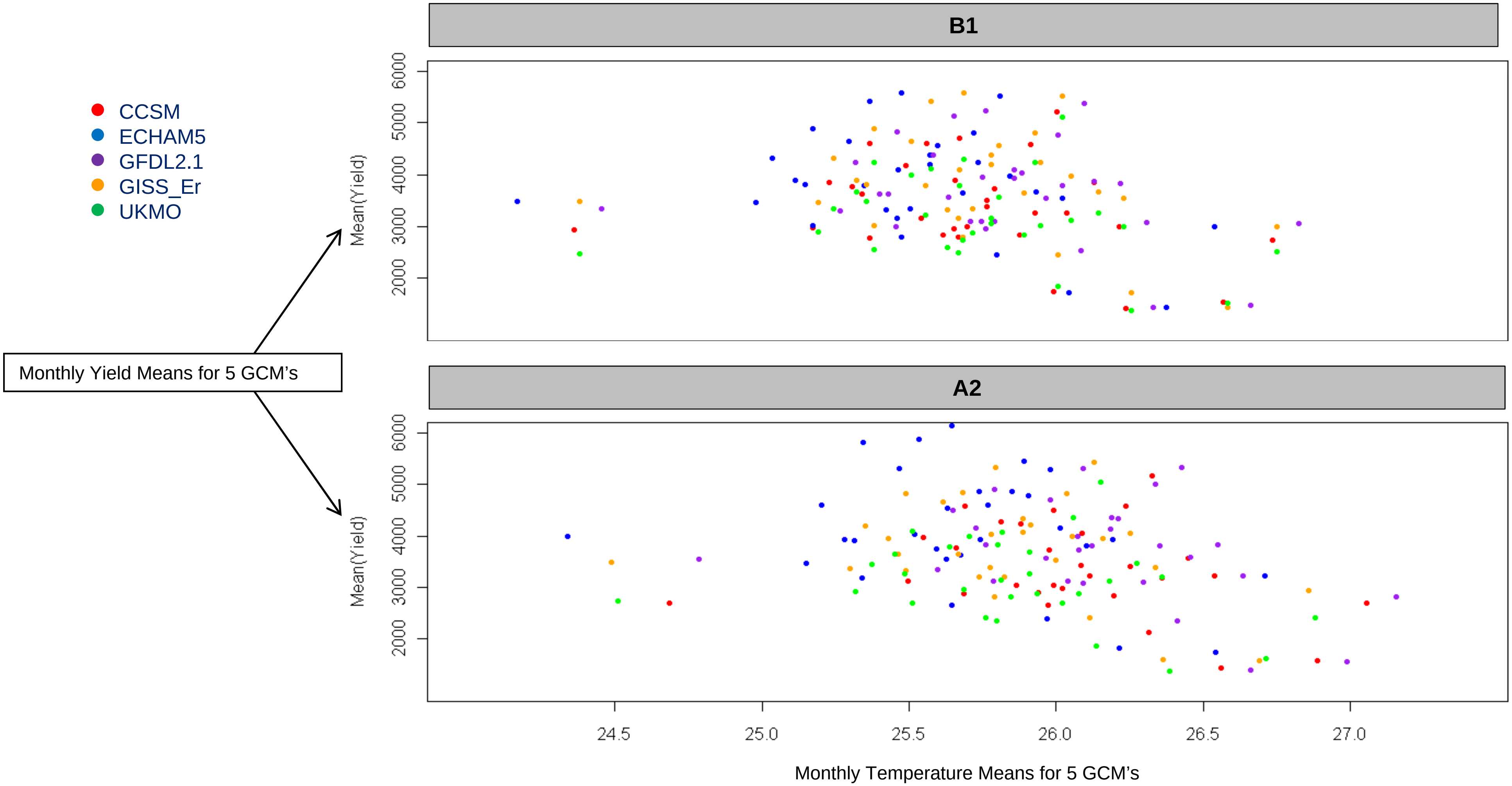
Average annual yield for GCM

NASA | USGS science for a changing world | Booz | Allen | Hamilton

Rainfall & GCM-Projected (2020-2049) Maize Yield By Scenario



Temp & GCM-Projected (2020-2049) Maize Yield By Scenario



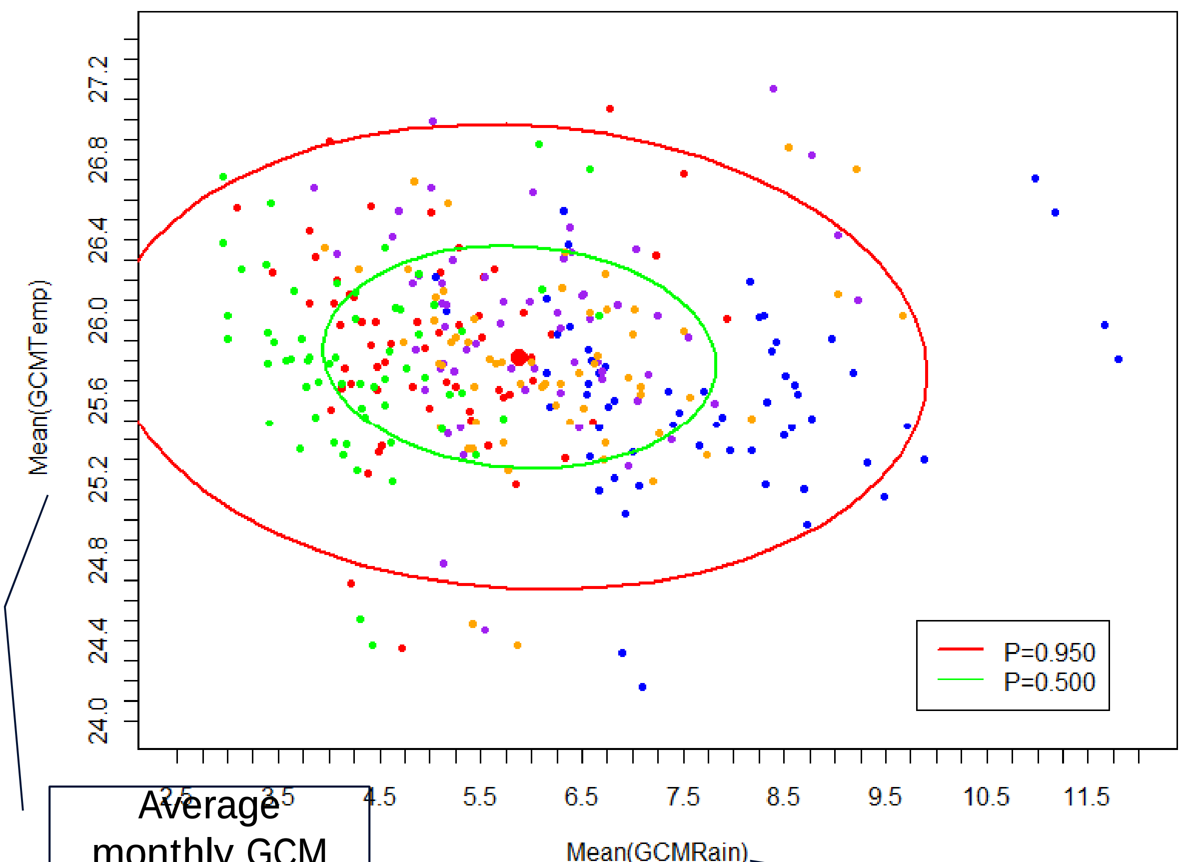
Booz | Allen | Hamilton

USGS science for a changing world

NASA

GCM Projections In Tocumen (2020-2049) For Temperature & Rainfall Display A Lack Of Correlation

Bivariate Fit of Mean(GCMTemp) By Mean(GCMRain)
Bivariate Normal Ellipses



Average monthly GCM temperature

Average monthly GCM rainfall

- ❖ **Temperature and rainfall are not significantly correlated (Pearson test)**
- ❖ **95% of the time**
  - **Temperature is between 24.6 and 26.6 Celsius**
  - **Rainfall is between 2.5 and 9.75 mm**
- ❖ **50% of the time**
  - **Temperature is between 25.3 and 26.4 Celsius**
  - **Rainfall is between 4 and 7.75 mm**

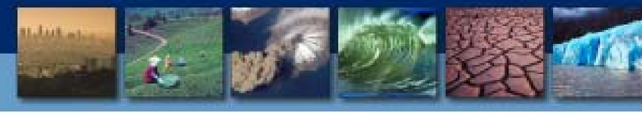
- CCSM
- ECHAM5
- GFDL2.1
- GISS_Er
- UKMO

NASA USGS science for a changing world Booz | Allen | Hamilton

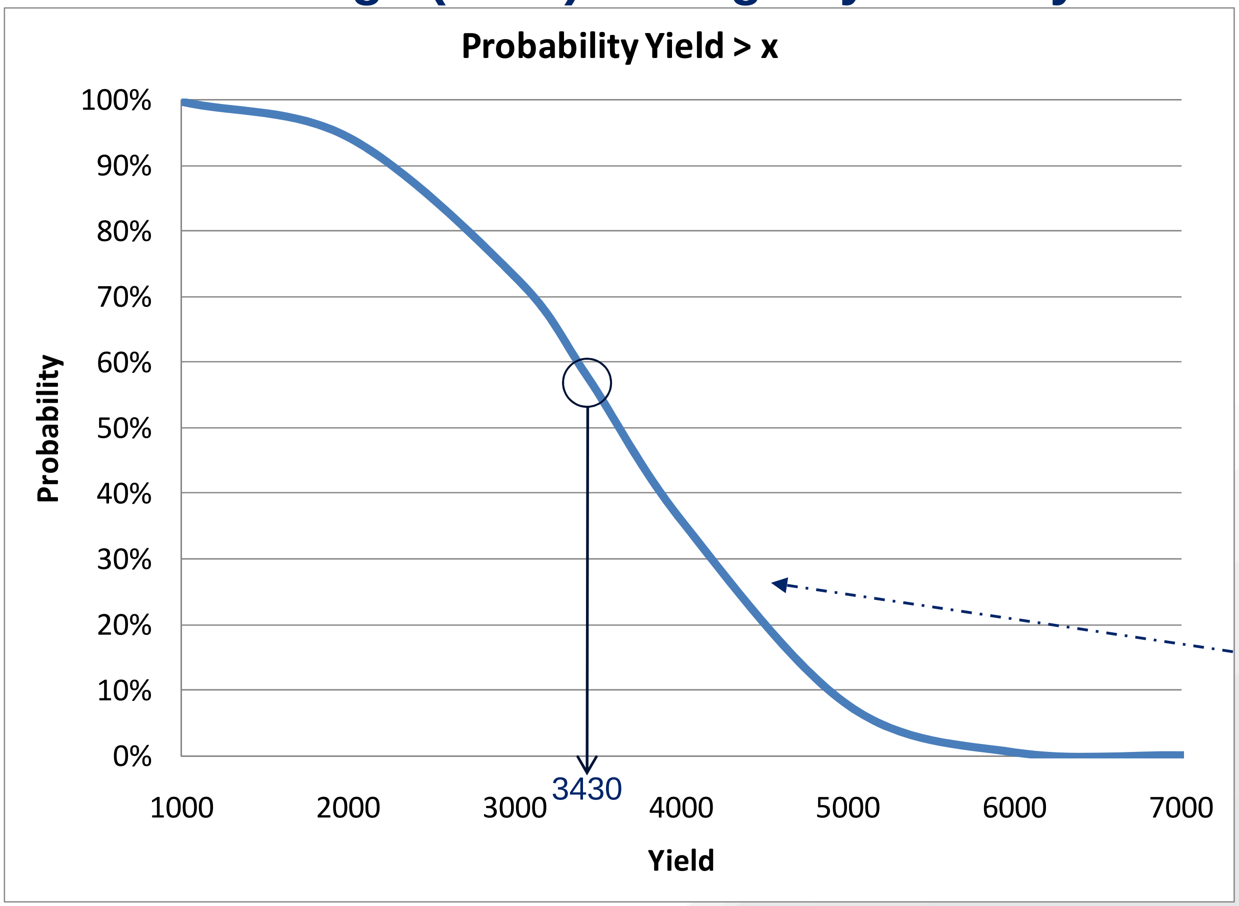
Outline

- ❖Sources Of Uncertainty in Climate Science and Applications
- ❖Crop Yield Prediction Process
- ❖ Analysis Of Results
- ❖ Uncertainty Characterization and Communication

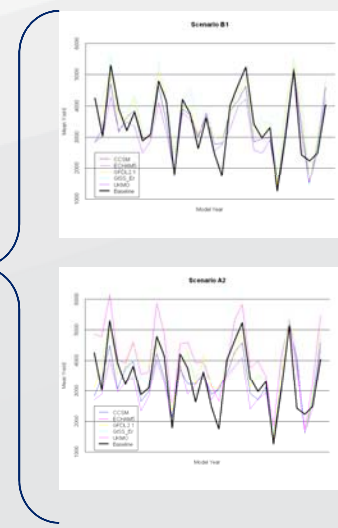
Distribution Analysis Deals With Uncertainty Across Years (2020-2049), GCMs, & Scenarios



❖ **The probability that Maize Yield will be above or below the baseline average (3430) during any future year**



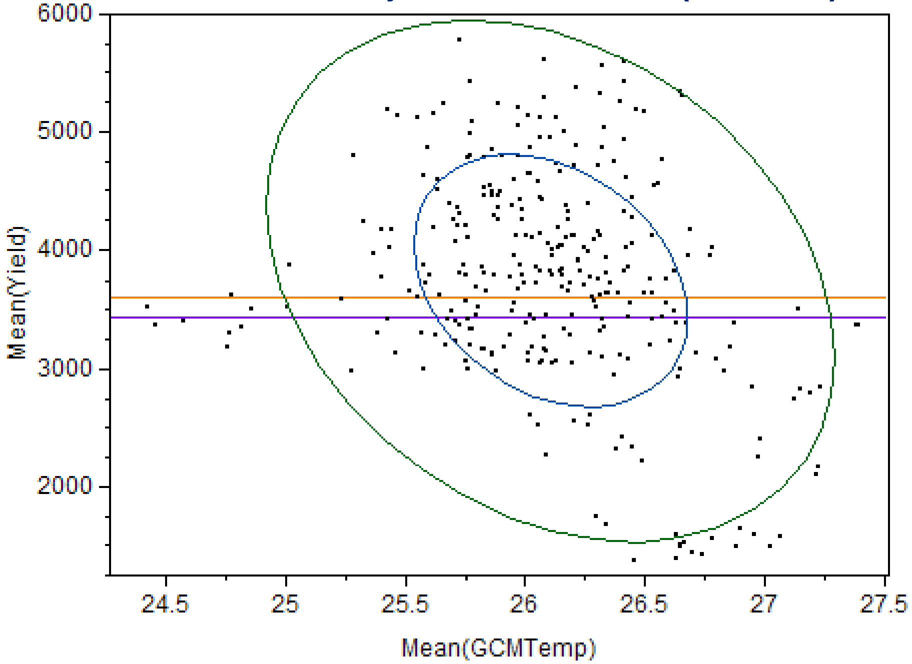
	Prob
Below Baseline (3430)	42%
Around Baseline (+/-10%)	26%
Above Baseline (3430)	58%



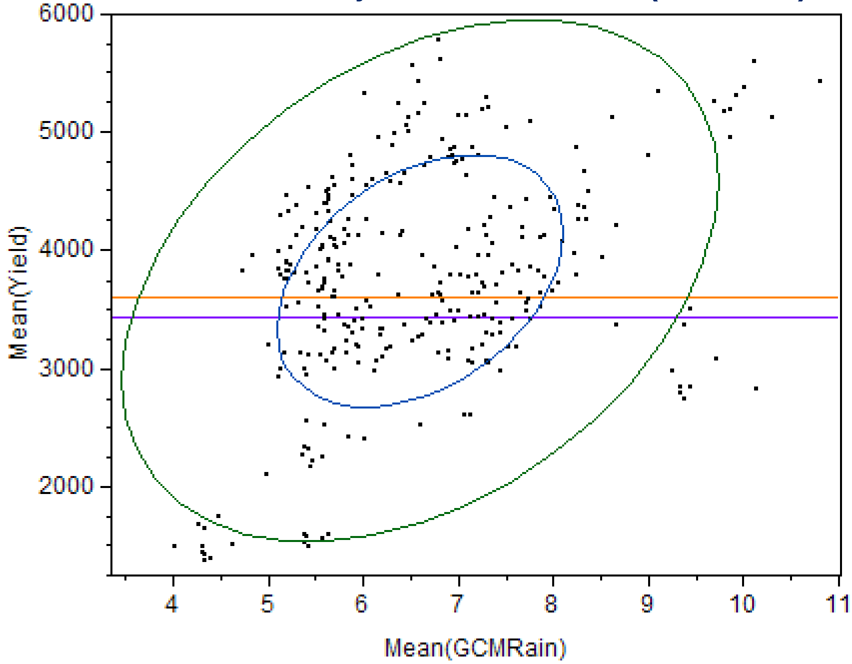
NASA USGS science for a changing world Booz | Allen | Hamilton

Bivariate Analysis Shows That Rainfall & Temp Do Affect GCM-Projected (2020-2049) Maize Yield

Temperature & Yield Percentages for CO2 GCM Projections in Tocumen (2020-2049)



Rainfall & Yield Percentages for CO2 GCM Projections in Tocumen (2020-2049)

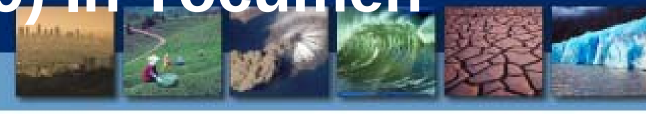


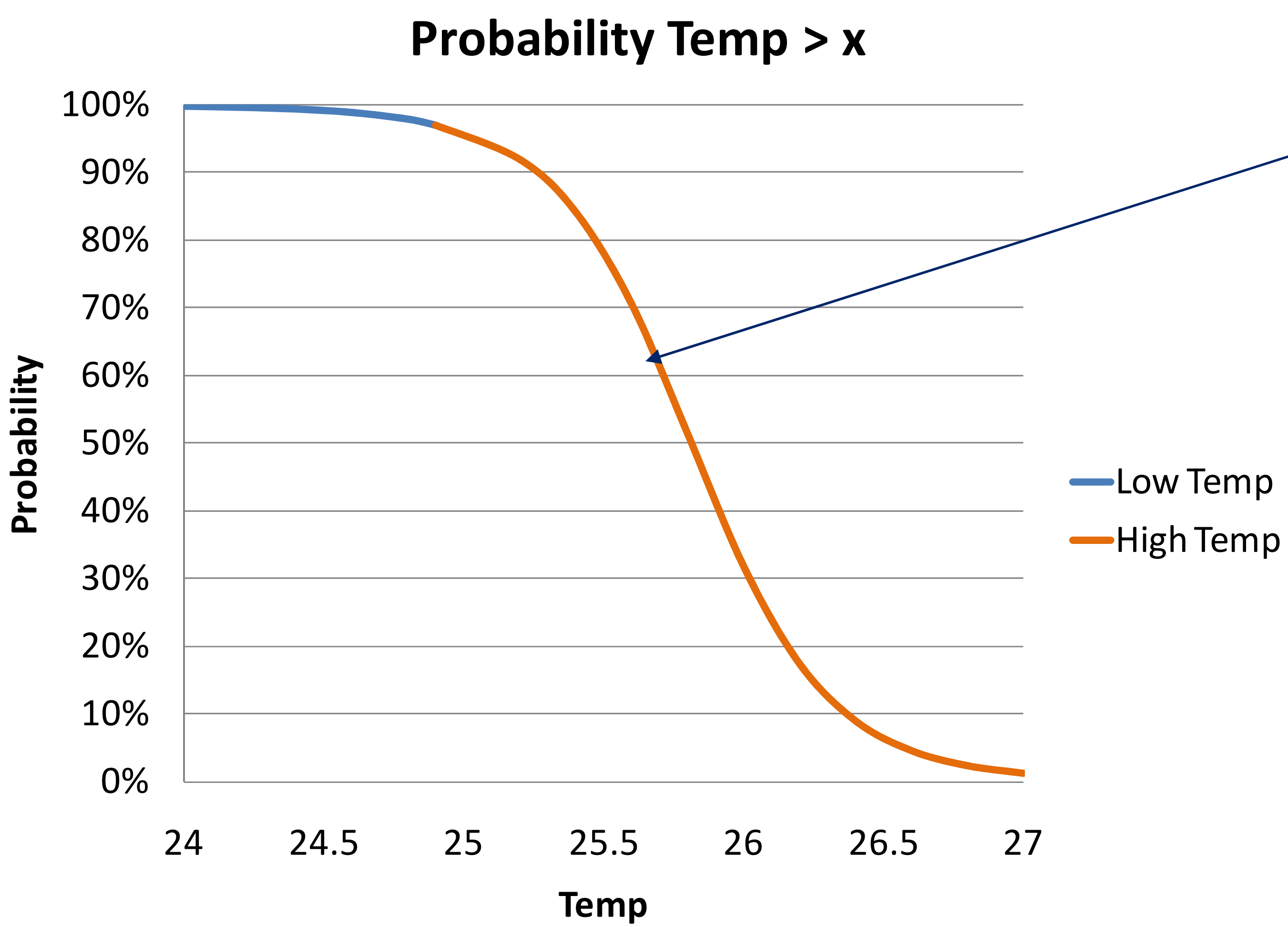
- ❖ **Follow-on analysis can determine**
  1. **optimal levels of temp and rain for maize yield**
  2. **the probability of encountering optimal conditions**

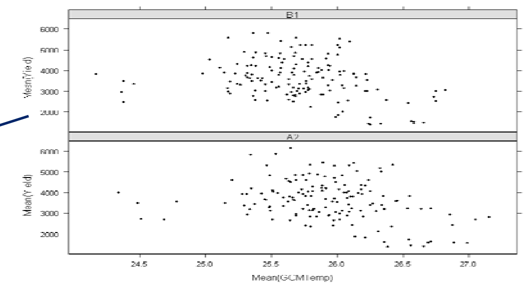
P = 0.950 Density Ellipse
P = 0.500 Density Ellipse

Baseline Average Yield = 3430
GCM Average Yield = 3607

Distribution Analysis Shows That GCMs Predict Hot Temperatures During Future Years (2020-2049) In Tocumen



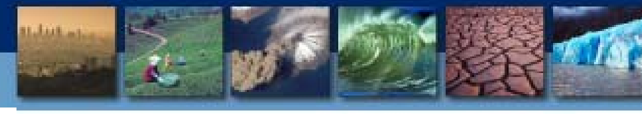


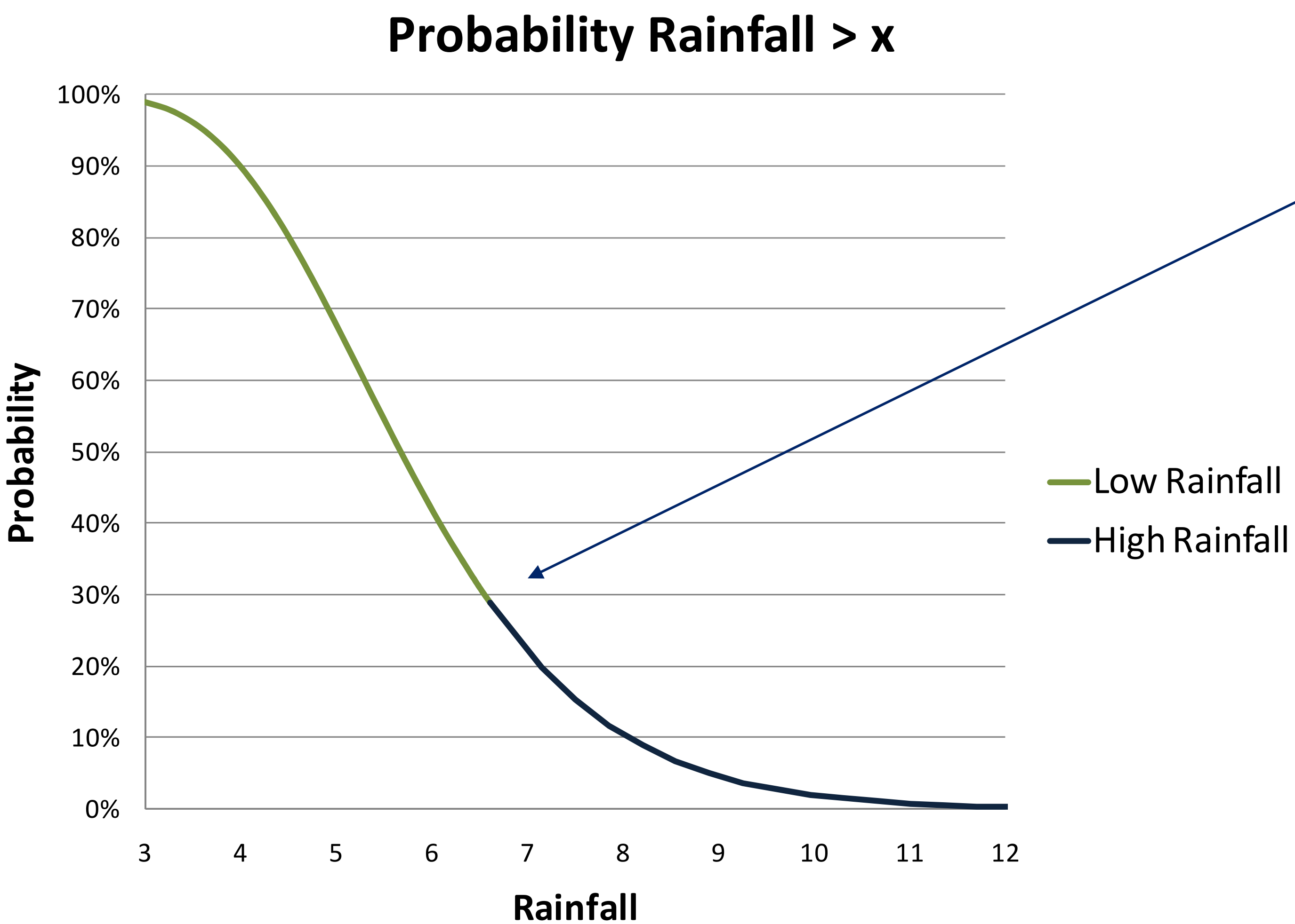


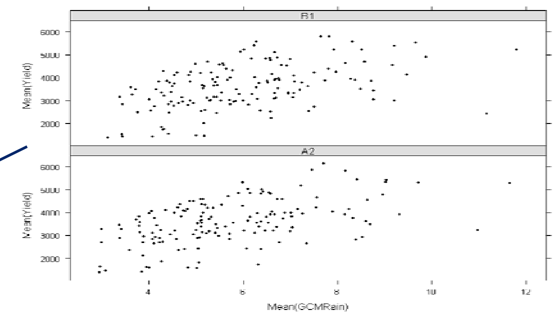
	Prb
Low Temp <24.9	3%
High Temp >24.9	97%

NASA USGS science for a changing world Booz | Allen | Hamilton

GCMs Predict That There Is An Almost Equal Probability Of Low Rainfall Vs High Rainfall During Future Years (2020-2049) In Tocumen







	Prb
Low Rain <6.6	71%
High Rain >6.6	29%

Rainfall & Temperature Far From Baseline Averages Impact The Probability Of Maize Yield For Any Future Year (2020-2049)

Low Temp, Low Rain	Prb
Below Baseline (3430)	79.74%
Around Baseline (+/-10%)	53.72%
Above Baseline (3430)	20.26%

Low Temp, High Rain

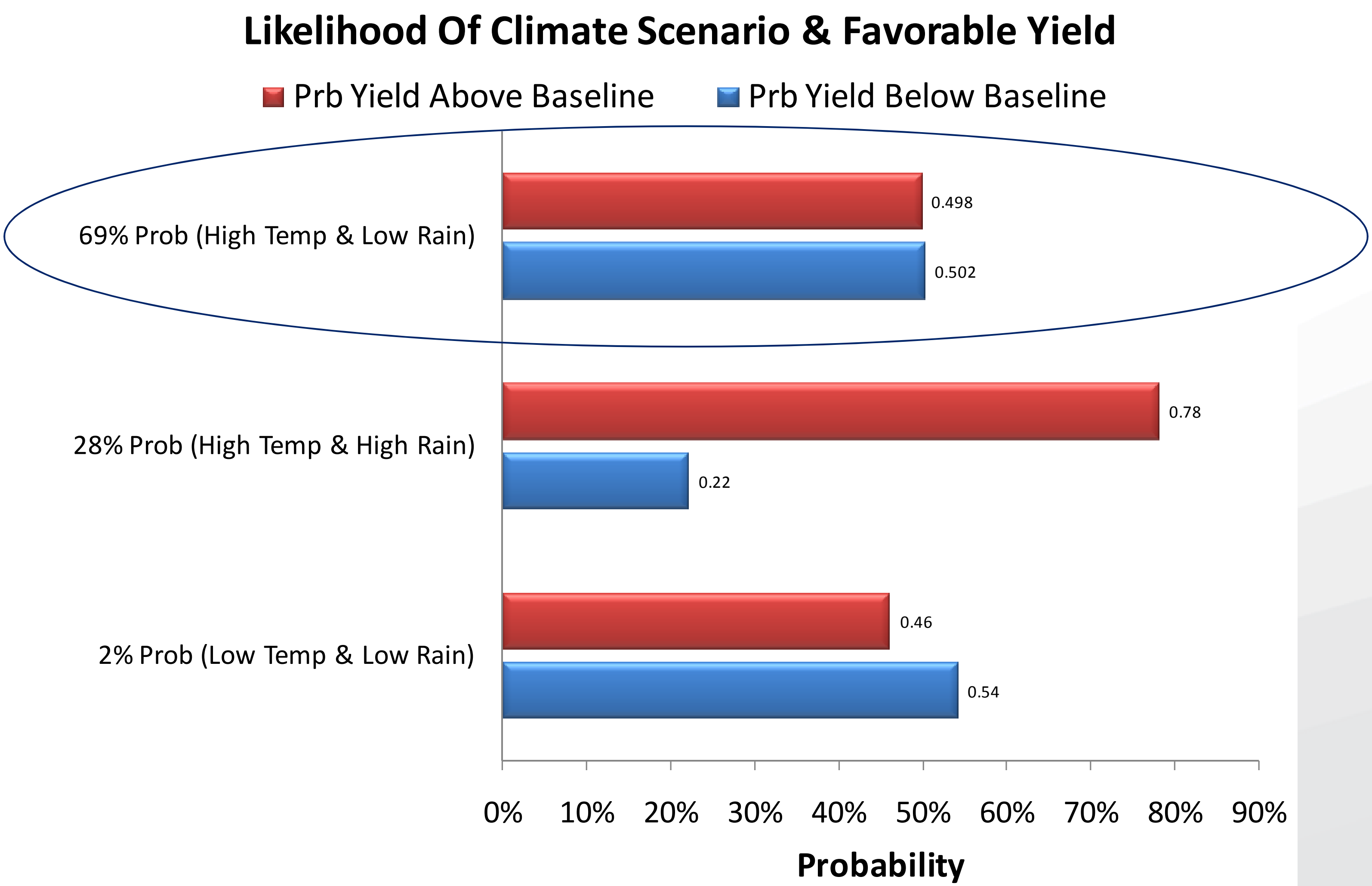
- Results are preliminary only
- Insufficient data available in this space
- A follow-on activity would be to model predictions through this space

High Temp, Low Rain	Prb
Below Baseline (3430)	50.20%
Around Baseline (+/-10%)	27.85%
Above Baseline (3430)	49.80%

High Temp, High Rain	Prb
Below Baseline (3430)	22.40%
Around Baseline (+/-10%)	19.82%
Above Baseline (3430)	77.60%

- Low Temp is defined as temp below the baseline average (24.9)
- Low Rainfall is defined as rainfall below the baseline average (6.6)

Knowing The Probability Distribution Of Temp & Rain Determines The Likelihood Of Maize Yield Scenarios (2020-2049)



Potential enhancement of decision support systems through better understanding of climate observational uncertainties and their impacts

- Development and validation of innovative integrated decision support tools
- Assessment of the success of data policies in terms of ensuring local data input to Decision Support System for Agrotechnology Transfer (DSSAT) and Mesoamerican Environmental Monitoring System (SERVIR)
- Assessment of the needs and opportunities for capacity building in Mesoamerica (e.g., which groups, subject matter, etc)
- Identification of areas where new scientific understanding is needed to improve products, assimilation systems and models, and decision support tools
- Characterize the likelihood of extreme values in factors such as rainfall and temperature, e.g. El Nino. Explore the consequences of extreme values on maize yield.
- Determine the optimal temperature range and optimal rainfall range for maize yield in a specific region
- Develop a probabilistic relationship between temperature, rainfall, and maize yield. This relationship will characterize the interaction between temperature and rainfall on maize yield. This is an improvement on the main-effects model previously built.
- Acquire raw satellite data and evaluate uncertainty in satellite measurements.

Developing a Measurement System Uncertainty Framework for Earth Observing Satellites

Nipa Phojanamongkolkij
NASA Langley Research Center
04/05/2011

The CLARREO mission

- The Climate Absolute Radiance and Refractivity Observatory (CLARREO) is an Earth Observing Satellite mission to provide accurate measurements to substantially improve understanding of climate change.
- CLARREO will include a Reflected Solar (RS) Suite, an Infrared (IR) Suite, and a Global Navigation Satellite System–Radio Occultation (GNSS-RO).
- CLARREO is in the mission formulation phase (Pre-Phase A.)

CLARREO Uncertainty Framework

- The uncertainty framework is a pre-planning study for CLARREO.
- Iterative one-year process involves
  - Interviewing and collaborating with formulation manager, project scientist, mission scientist, IR instrument scientist, calibration manager, and research physical scientists.
  - Reviewing mission related materials.

CLARREO Uncertainty Framework

Mission Goals

MG-P Prove the ability to make global and zonal measurements with the accuracy, sampling and information content necessary to:

MG-P1 Detect annual and decadal climate change trends and

MG-P2 Test and improve global climate model predictions.

CLARREO Uncertainty Framework

Mission Goals

MG-P Prove the ability to make global and zonal measurements with the accuracy, sampling and information content necessary to:

MG-P1 Detect annual and decadal climate change trends and

MG-P2 Test and improve global climate model predictions.

Reflected Solar (RS) | Infrared (IR) | GNSS-RO

Science Objectives

SO-P Decadal climate change observations of *forcings*: Total and Spectral Solar Irradiance. (NOAA: TSIS).

SO-P Decadal climate change observations of *responses*:
SO-P1 Temperature and humidity profiles.
SO-P2 Cloud properties.
SO-P3 Top of atmosphere shortwave and longwave radiative fluxes.

SO-P Decadal climate change observations of *feedbacks*:
SO-P4 Cloud feedback, largest uncertainty
SO-P5 Water vapor and lapse rate feedback

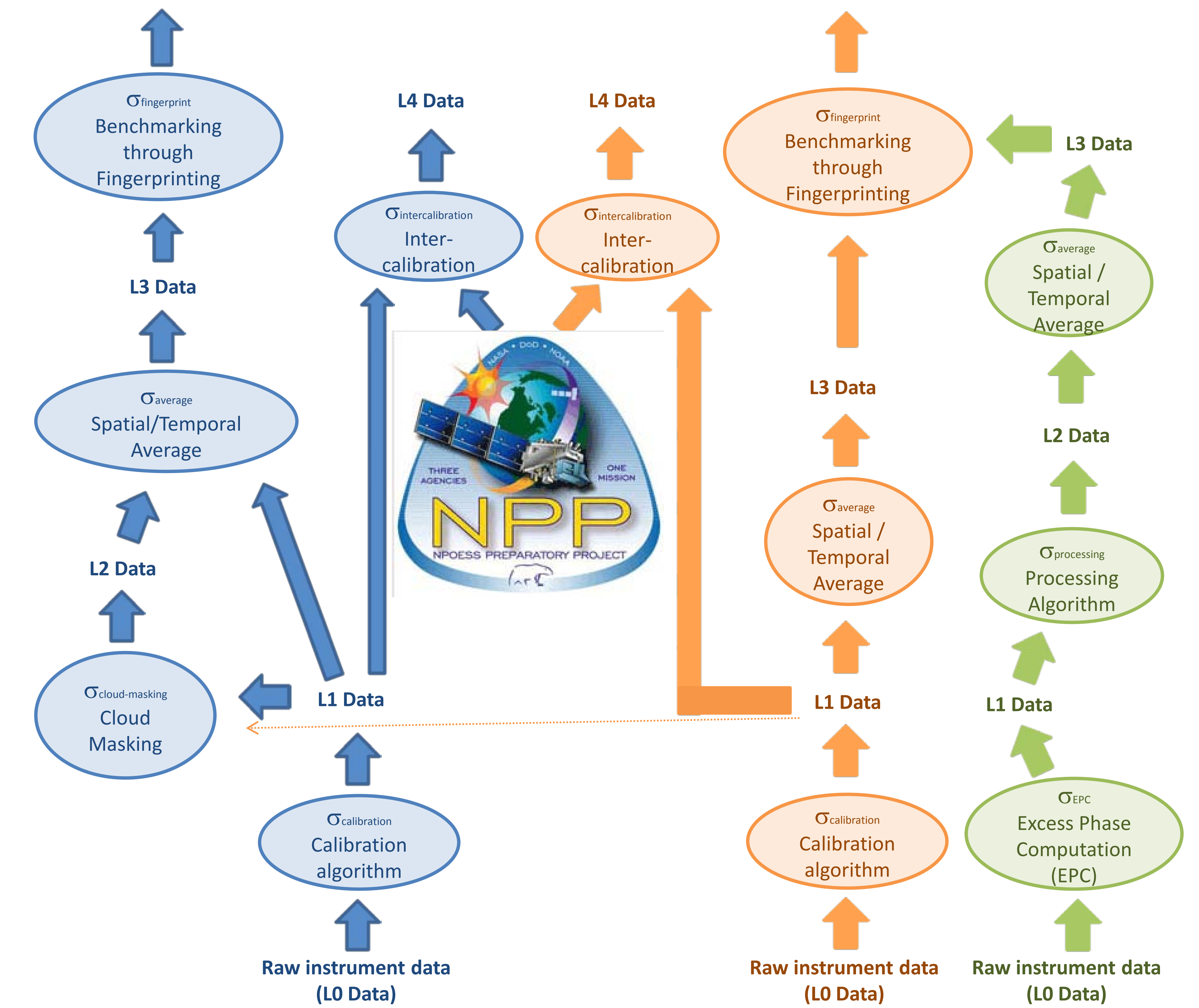
σ_{GCM} Test and improve global climate model (GCM)

σ_{trend} Detect annual and decadal trends

Time series of climate observations

Forcings:			
- Surface Albedo	RS		
- Greenhouse Gas		IR	
Feedbacks:			
- Cloud	RS	IR	
- Water Vapor		IR	GNSS-RO
- Lapse Rate		IR	GNSS-RO
- Snow/Ice Albedo	RS		

Responses:			
- Temperature profiles		IR	GNSS-RO
- Water Vapor profiles		IR	GNSS-RO
- Cloud response	RS	IR	
- Radiation response	RS	IR	
- Snow/Ice Cover	RS		



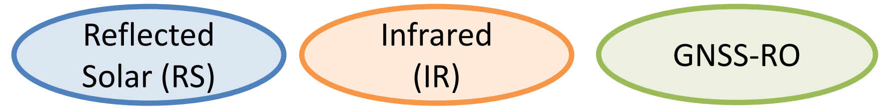
CLARREO Uncertainty Framework

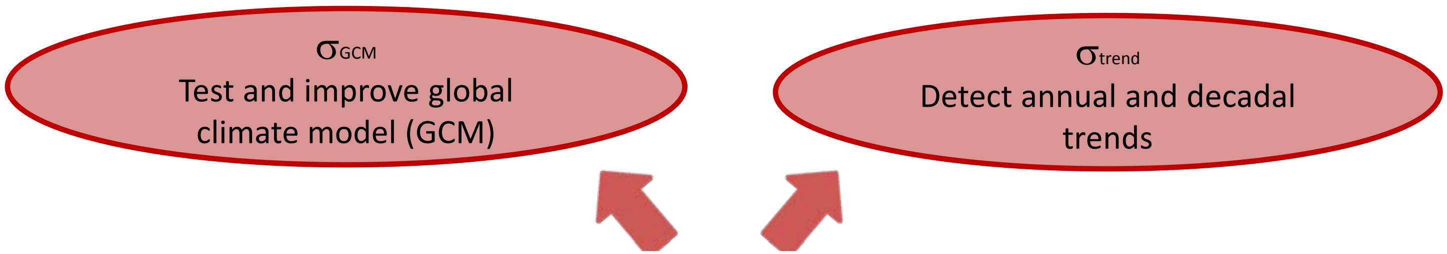
Mission Goals

MG-P Prove the ability to make global and zonal measurements with the accuracy, sampling and information content necessary to:

MG-P1 Detect annual and decadal climate change trends and

MG-P2 Test and improve global climate model predictions.

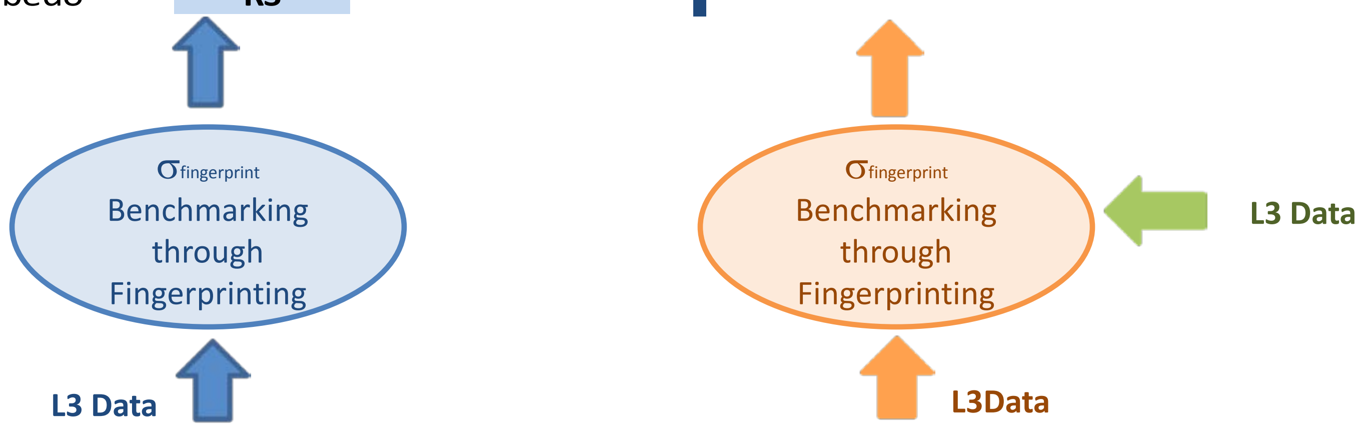




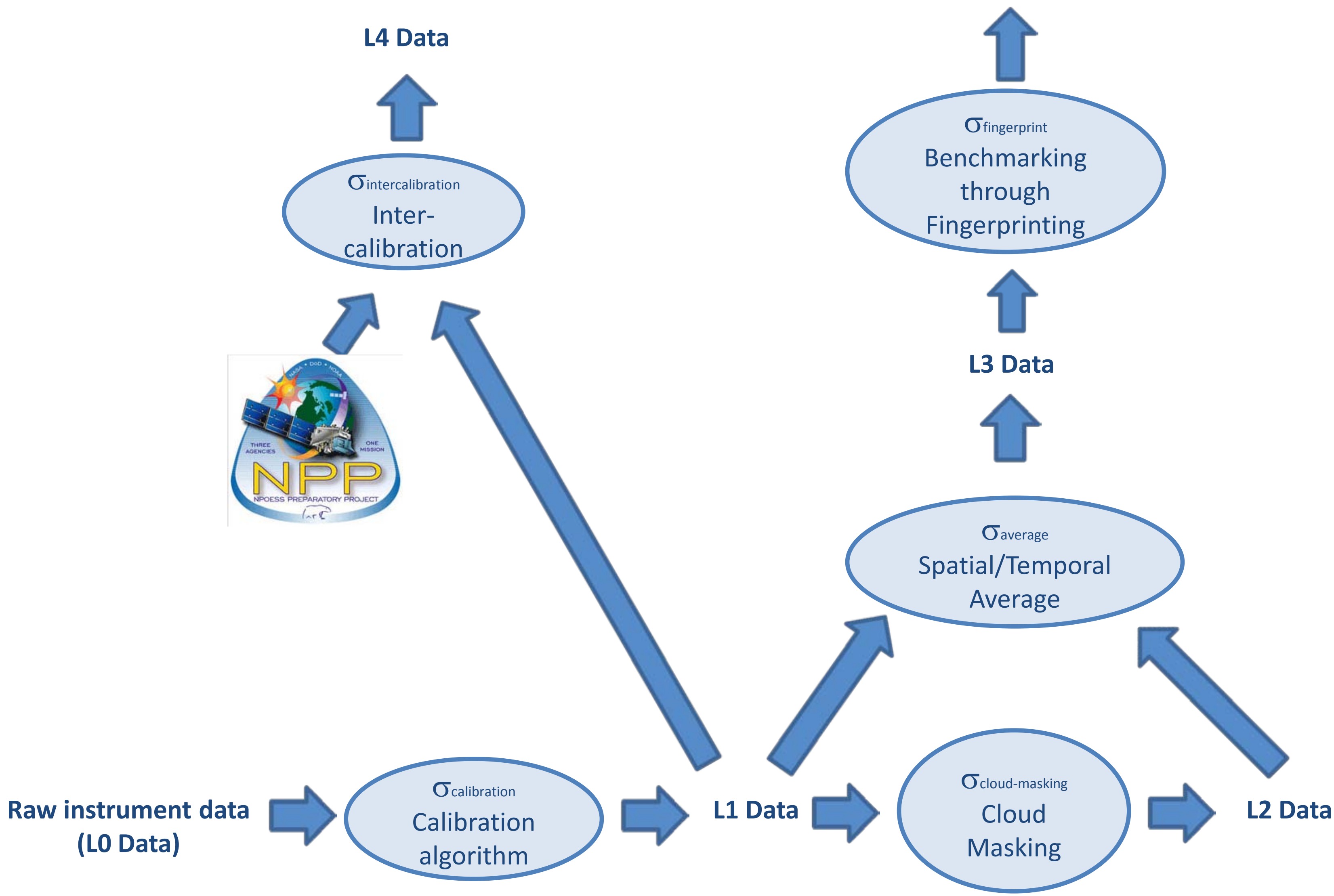
Time series of climate observations

Forcings:			
- Surface Albedo	**RS**		
- Greenhouse Gas		**IR**	
Feedbacks:			
- Cloud	**RS**	**IR**	
- Water Vapor		**IR**	**GNSS-RO**
- Lapse Rate		**IR**	**GNSS-RO**
- Snow/Ice Albedo	**RS**		

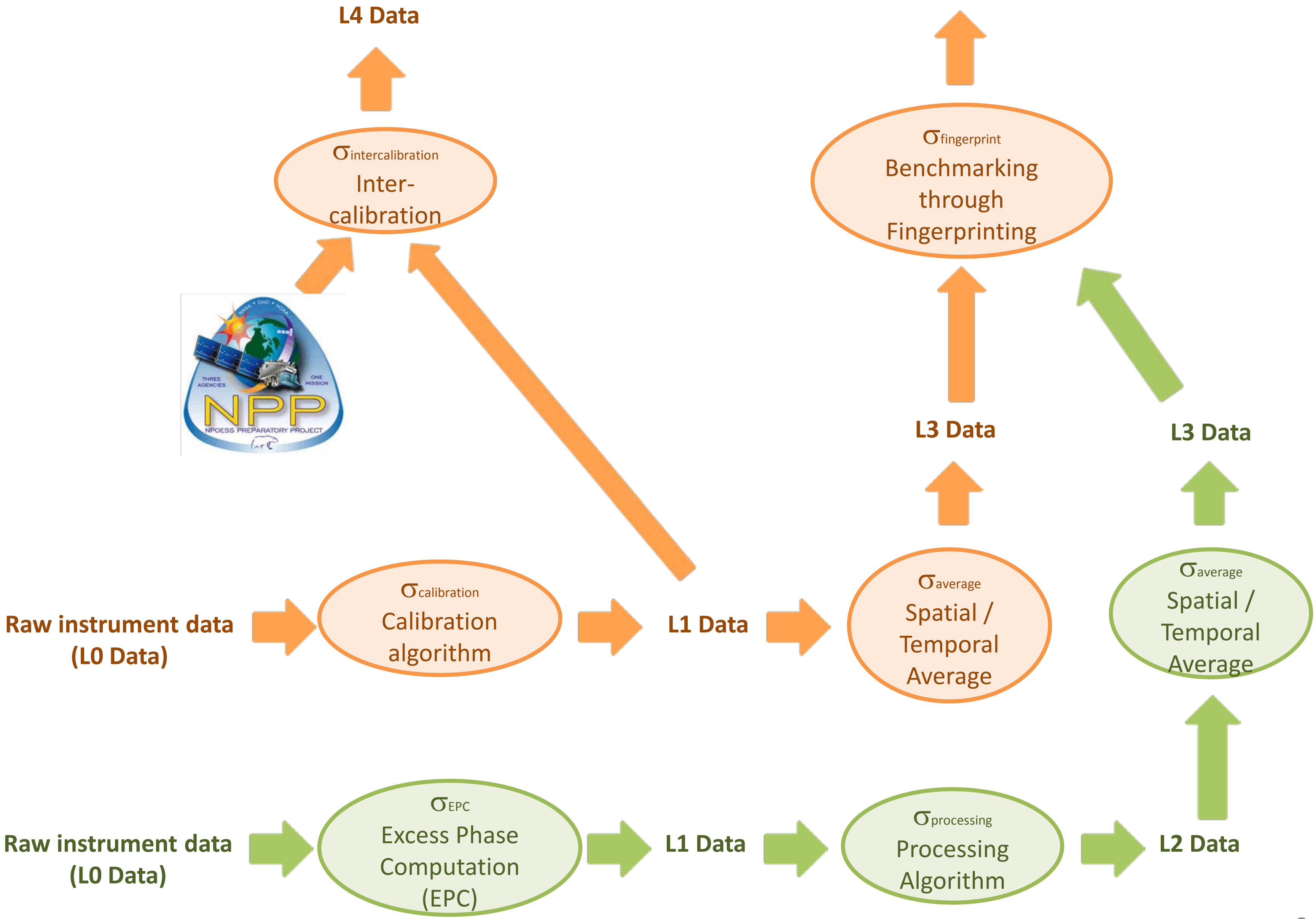
Responses:			
- Temperature profiles		**IR**	**GNSS-RO**
- Water Vapor profiles		**IR**	**GNSS-RO**
- Cloud response	**RS**	**IR**	
- Radiation response	**RS**	**IR**	
- Snow/Ice Cover	**RS**		



CLARREO Uncertainty Framework - RS



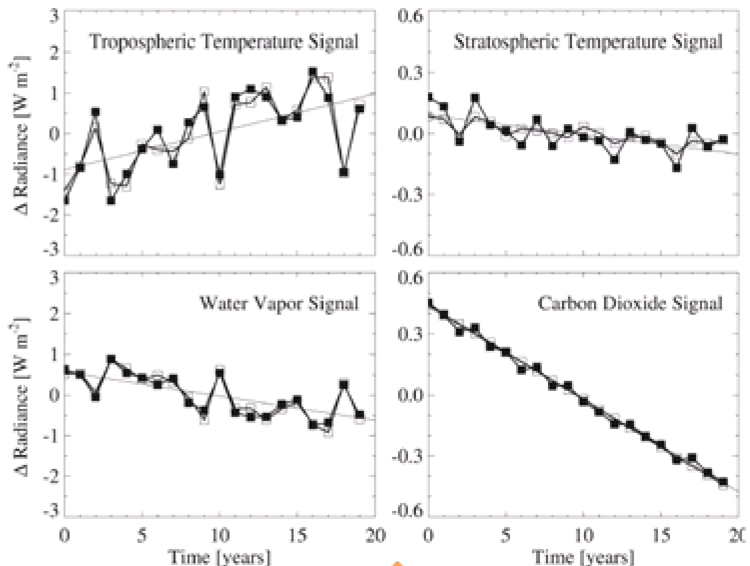
CLARREO Uncertainty Framework – IR and GNSS-RO



σ_{GCM}
Test and improve global climate model (GCM)

σ_{trend}
Detect annual and decadal trends

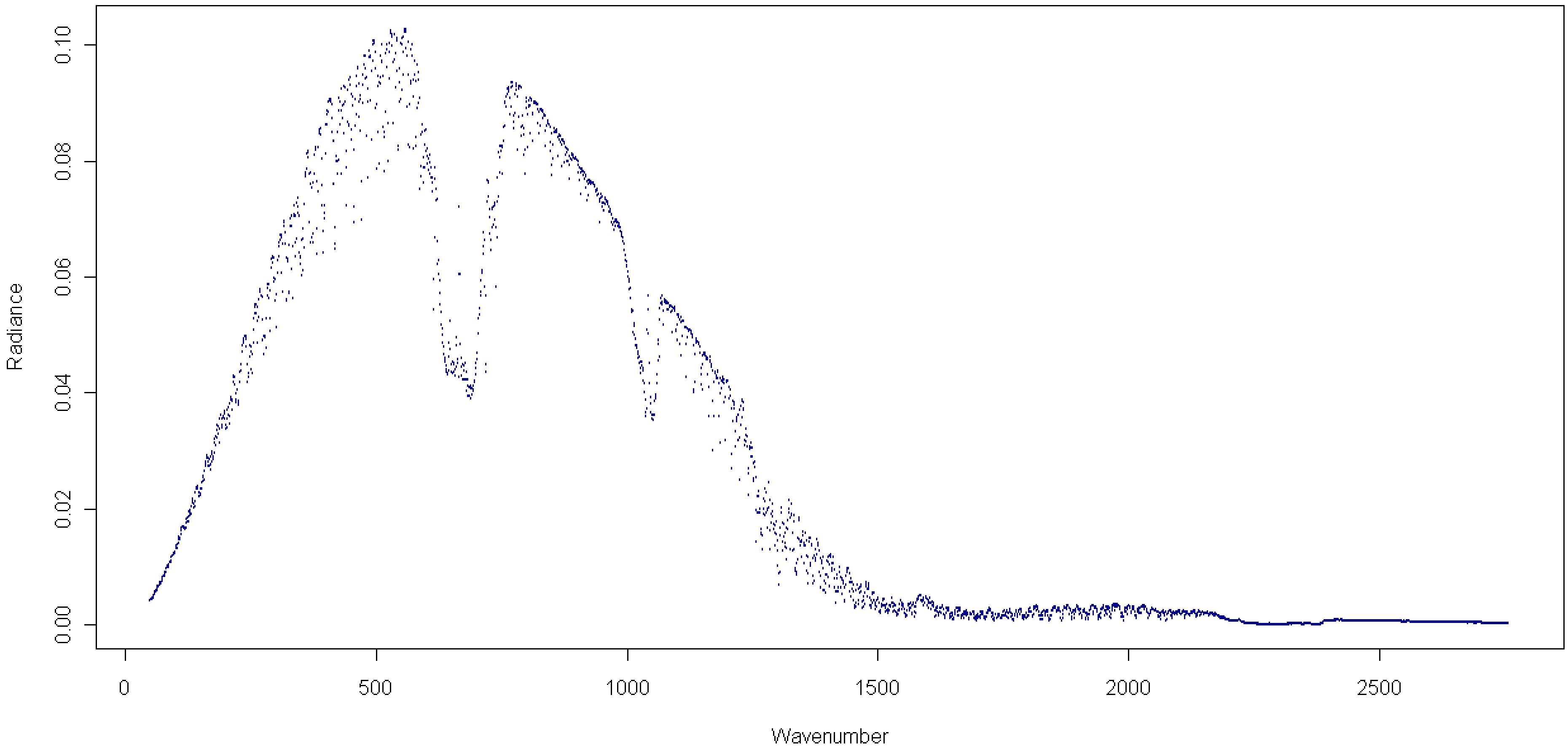
Time series of climate observations



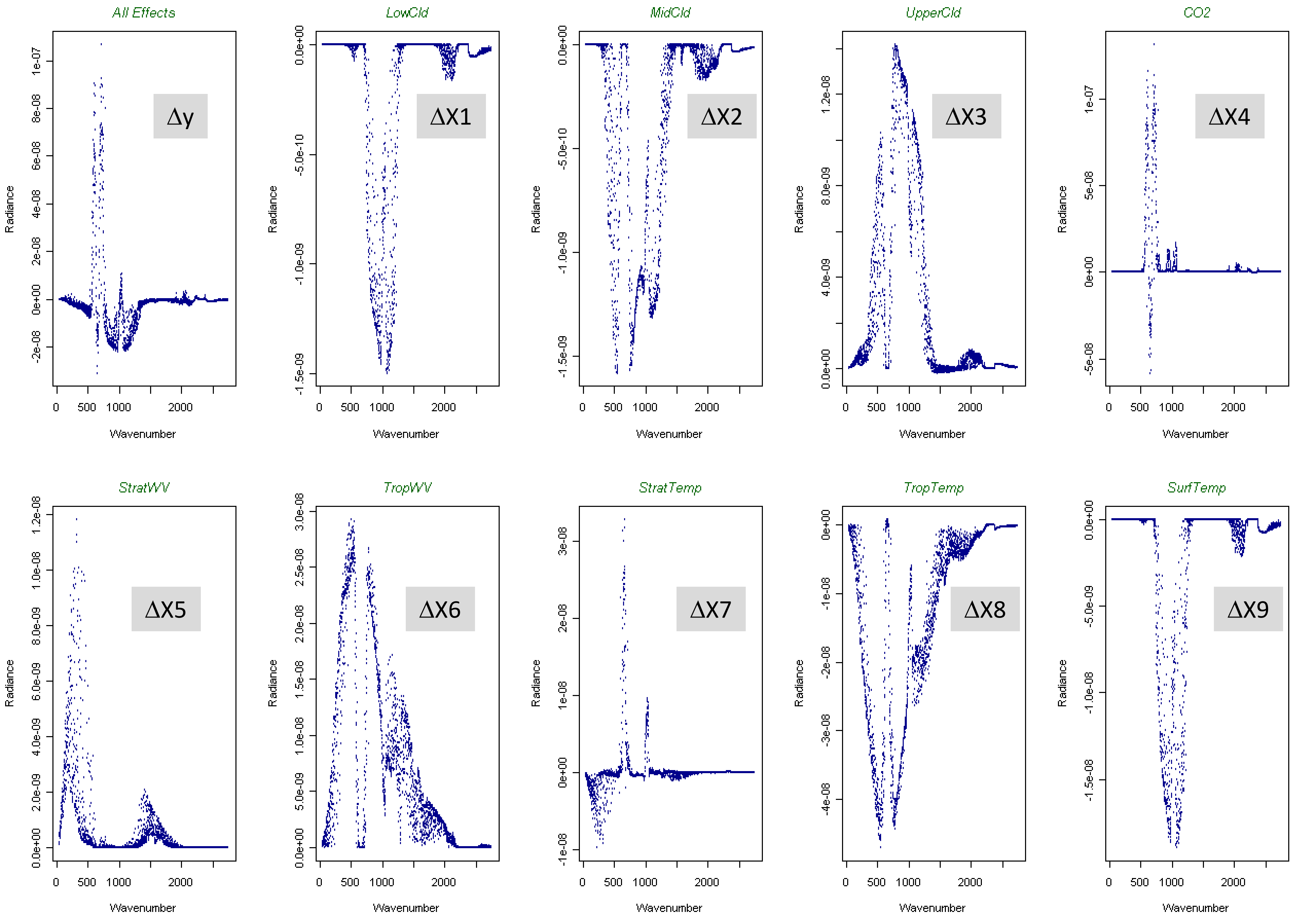
$\sigma_{fingerprint}$
Benchmarking through Fingerprinting

L3 Data

Calibrated Measurement (L3 data)

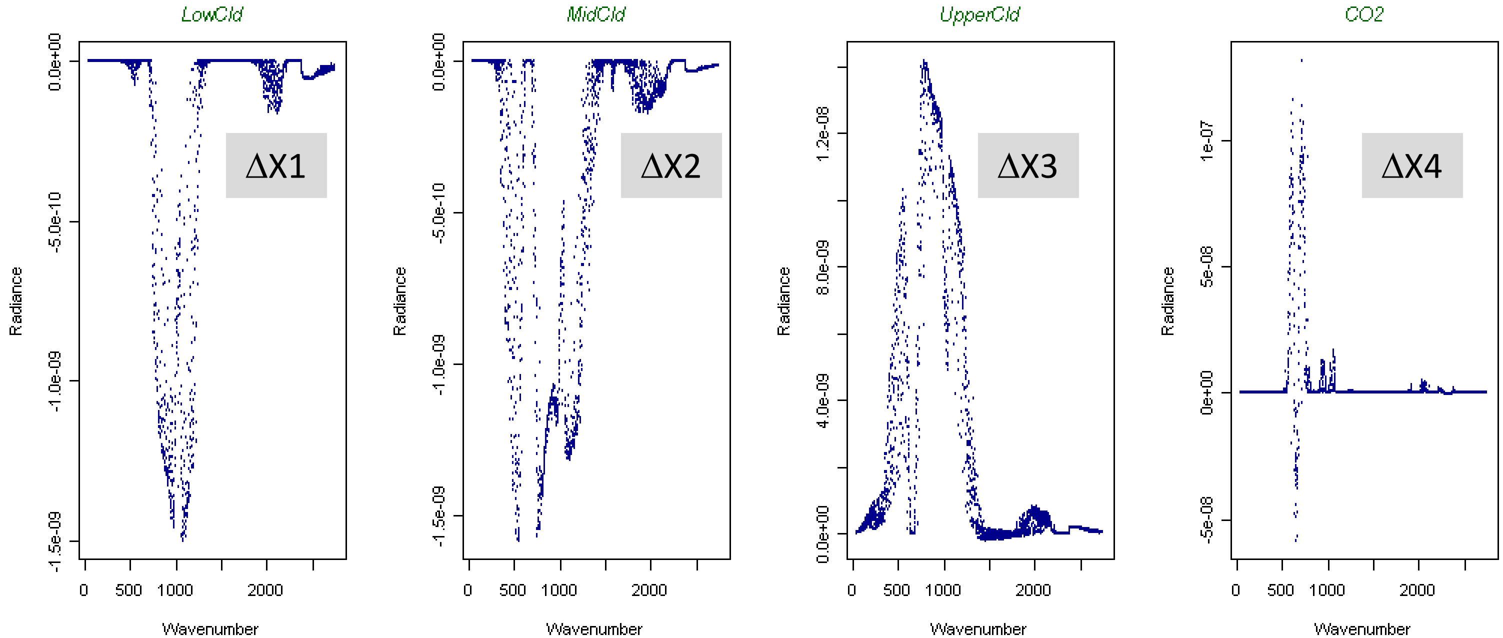


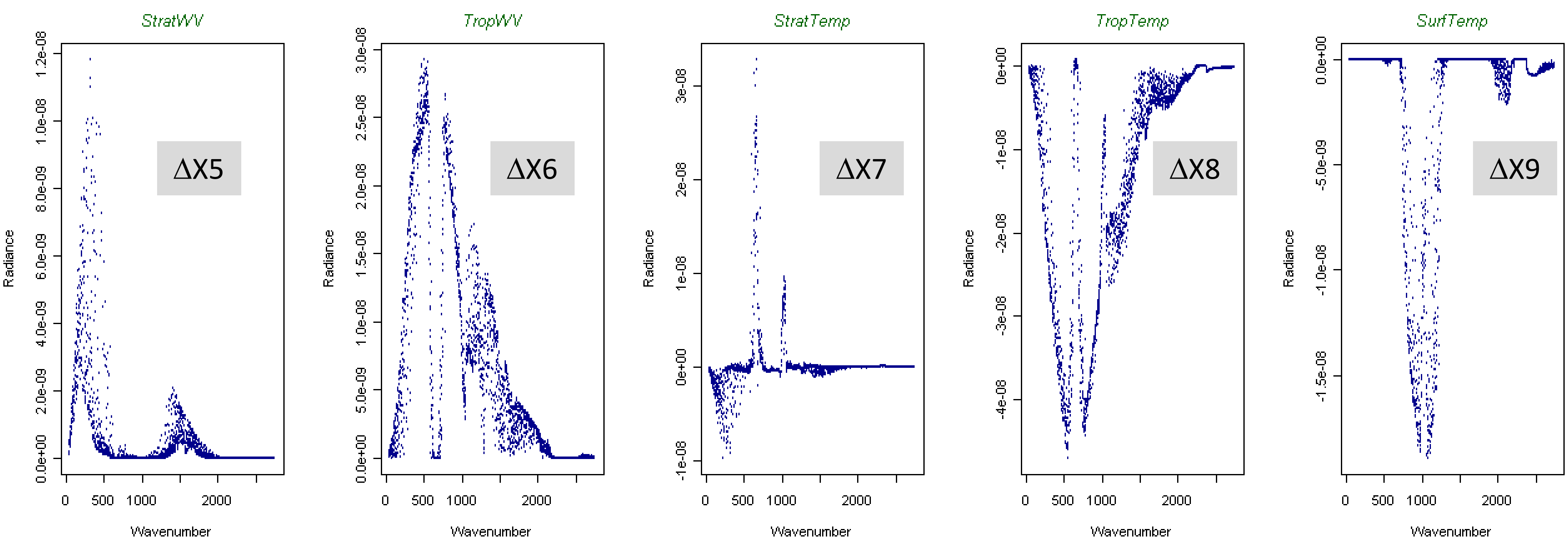
Signal Signatures – Spectral Correlation



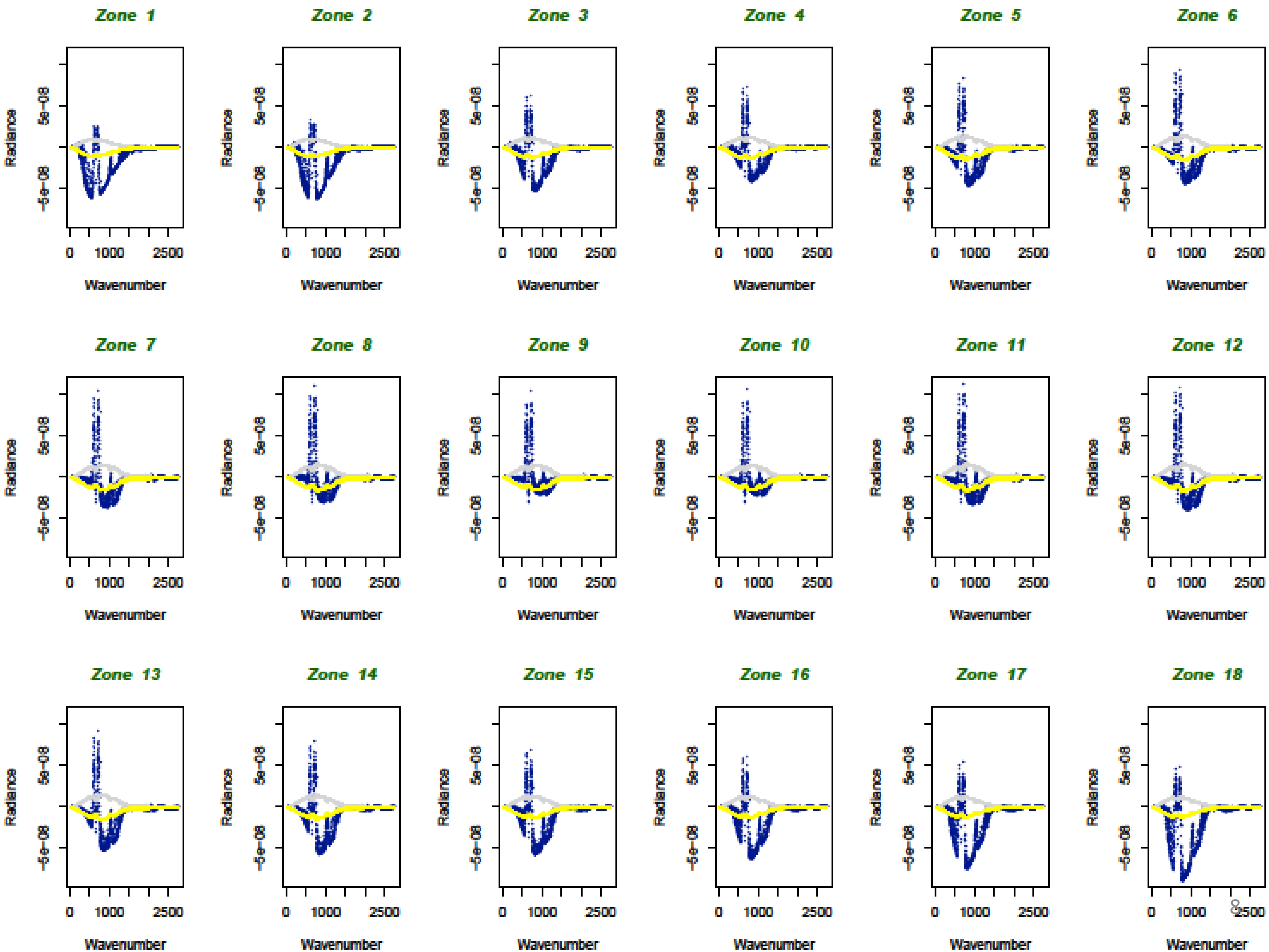
Signal Signatures – Multi-Collinearity

Similarity of Signatures

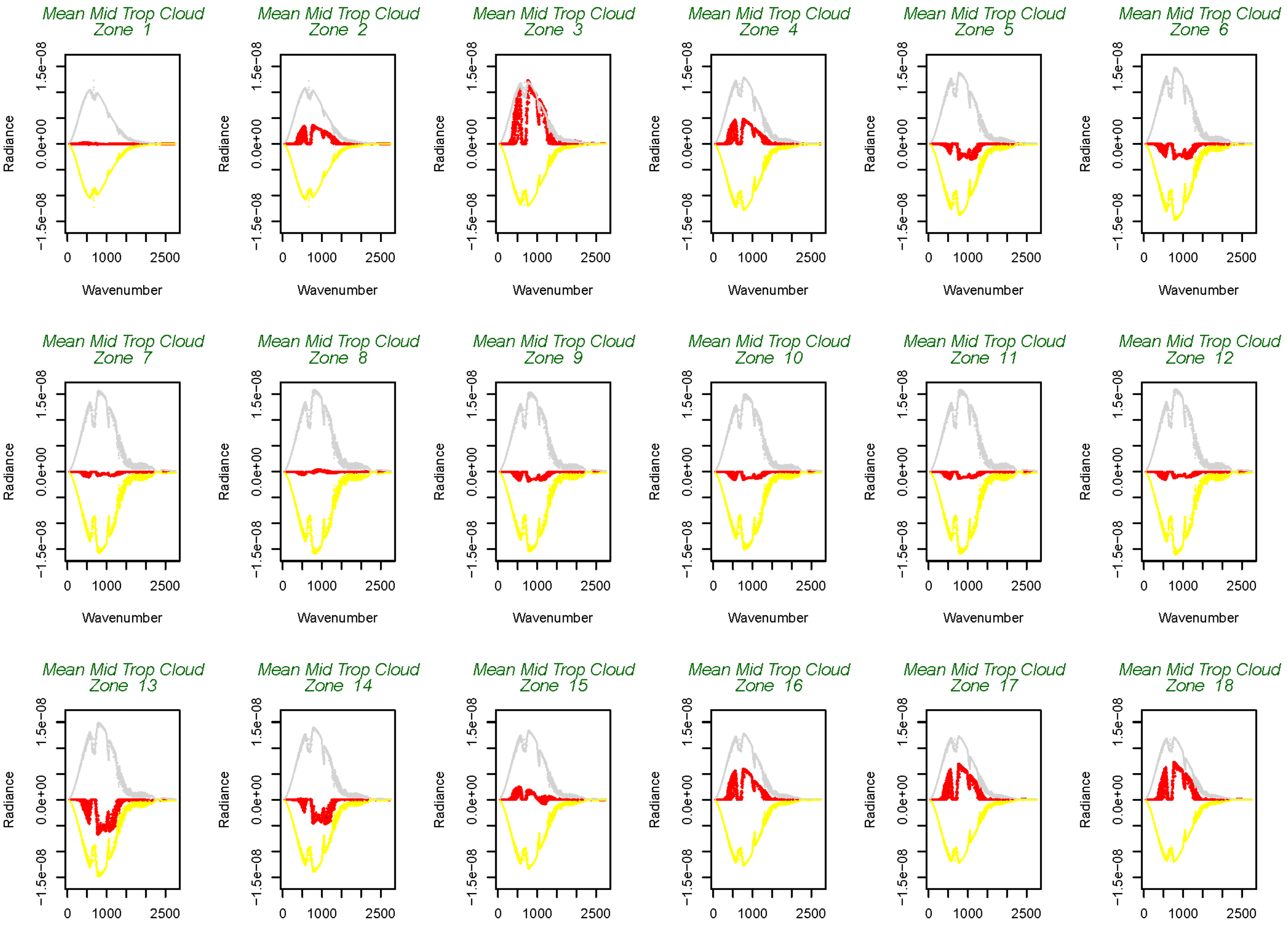




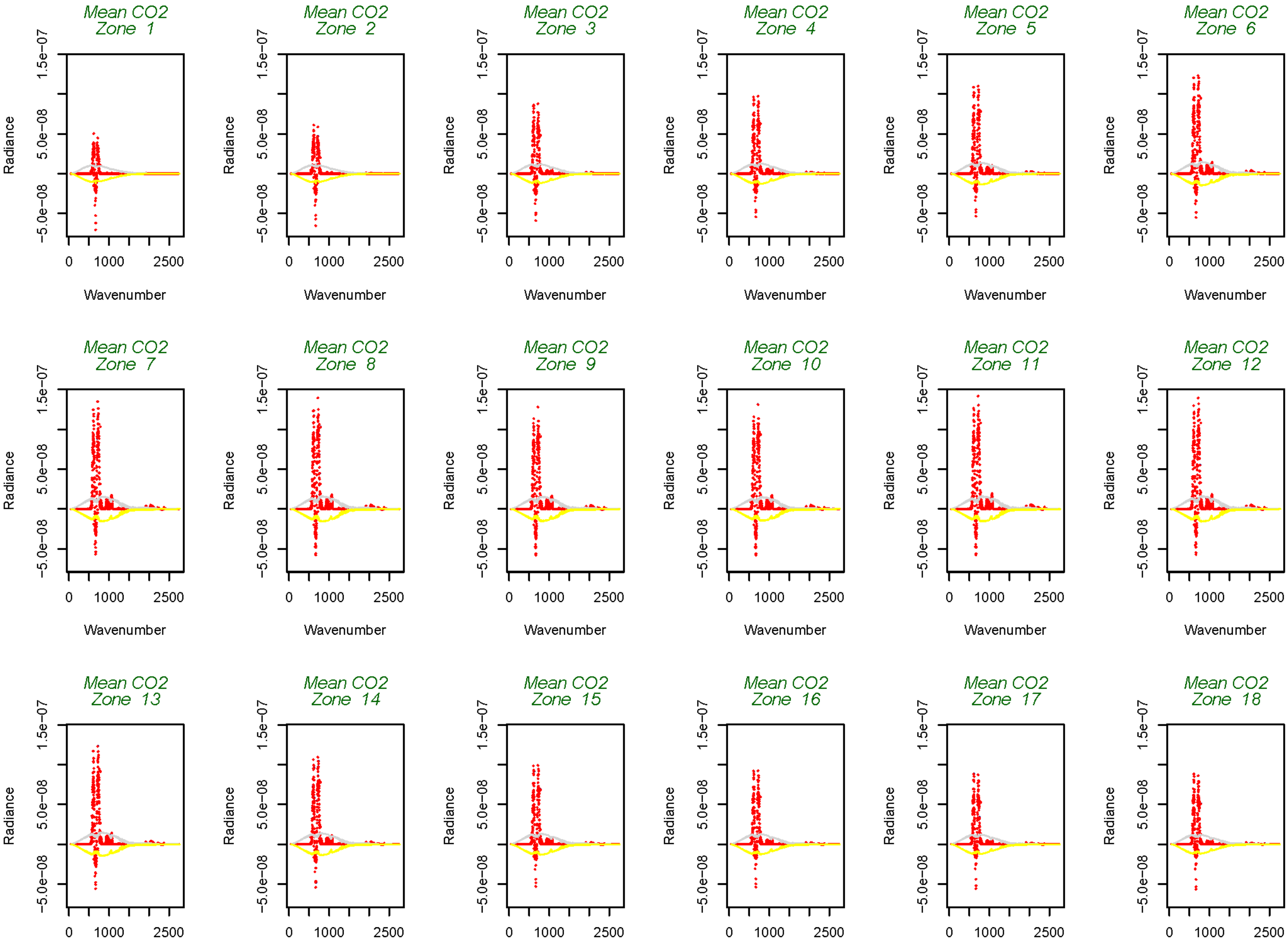
Measurement Radiance Difference (Δy) – Spatial Variation



ΔX2 – Spatial Variation



ΔX4 – Spatial Variation



Fingerprint Challenges

Multi-Collinearity

$$\Delta y = a_1 \Delta X_1 + \ldots + a_9 \Delta X_9 + \varepsilon$$

Spectral Correlation
Spatial Correlation

Spectral Correlation
Spatial Correlation

Spectral Correlation
Spatial Correlation

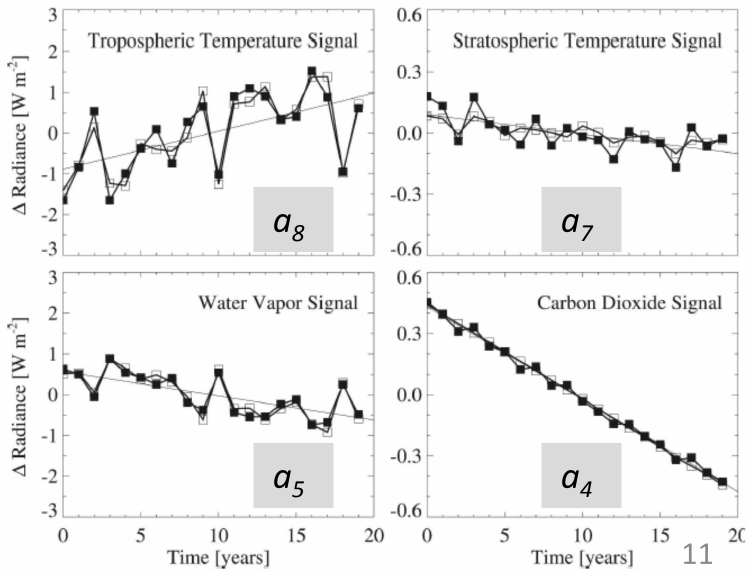


Figure 1 Time series samples obtained from fingerprinting analysis.

Conclusion

In developing a measurement system for Earth Observing Satellites to meet the accuracy level set by the mission to fulfill the goals, we need to

- Understand the end-to-end process from raw measurement to final science data.
- Evaluate the uncertainty budget for all steps of the end-to-end process.
- Determine the critical uncertainty driver(s) that could potentially affect the accuracy requirement.
- Allocate resources accordingly to these drivers.

Design For Variation

NASA Statistical Engineering Symposium

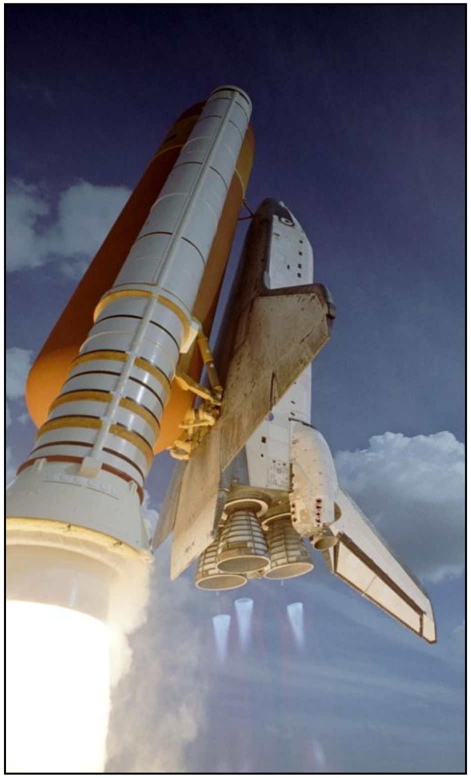
Williamsburg, VA 5/5/2011

Grant Reinman, Fellow, Statistics

Pratt & Whitney, East Hartford, CT

Pratt & Whitney Engineering

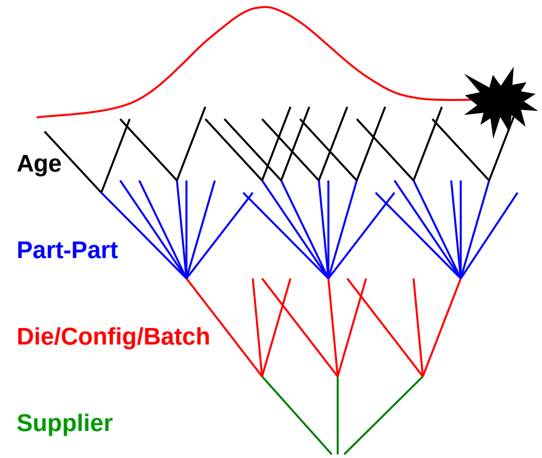
A Passion for innovation



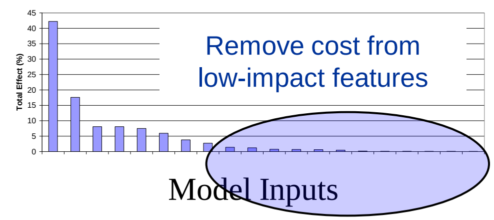
Design For Variation

A Strategic Initiative at Pratt & Whitney

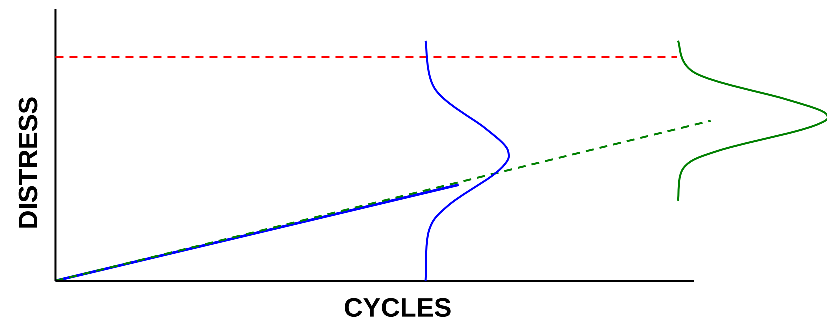
- ▲ To Reduce Escapes (Safety)
  - – Variation plays a significant role in field problems
  - – Cost of finding/correcting problems increases rapidly as product matures



- ▲ To Improve Producibility (Cost/Competitiveness)
  - – Find and focus on important features (few?)
  - – Relax requirements on unimportant features (many?)
  - – Use Robust Design to reduce sensitivity



- ▲ To Maximize Rotor Life (Time on Wing)
  - – Rotor life depends on max distress / min life airfoil
  - – 'Weakest link' structure pervasive in gas turbines
  - – Reducing variation increases rotor life



Design For Variation (DFV) Strategic Plan

Vision: All Key Modeling Processes will be DFV-enabled

- **Strategy**
  - ☑ Identify Key Processes
  - ☑ Define elements of a DFV-enabled modeling process
  - ☑ Provide Resources under Strategic Initiative

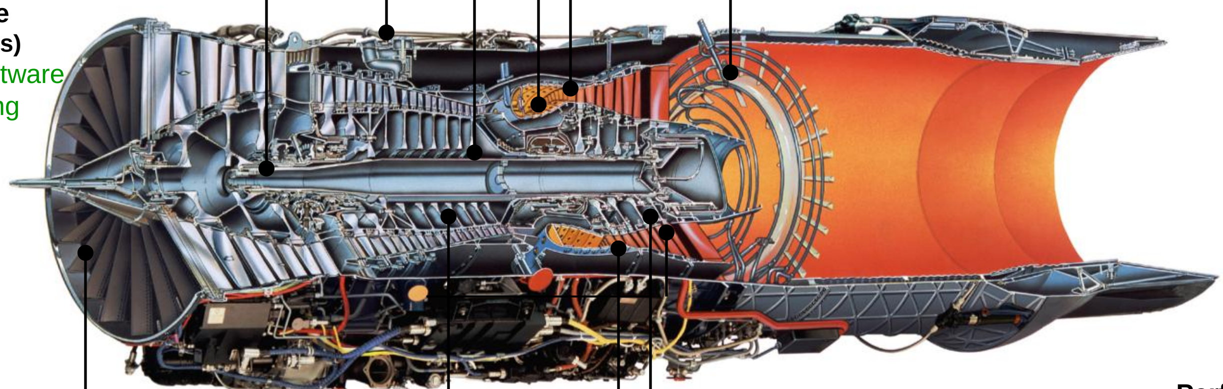
Black: Legacy Task
Green: 2010 funded
Blue: 2011 funded (new)

Mechanical Systems and Externals
Carbon Seal Performance
Ball & Roller Bearing Design
FDGS Durability
Externals: Forced Response Analysis

Combustor and Augmentor
Combustor pattern factor
Combustor Liner TMF
Augmentor Ignition Margin Audit
Mid Turbine Frame Robust Design

Air Systems
Thermal Management Model
Internal Air System Model
Engine Data Matching

DFV Infrastructure
(Statistics & Partners)
Sens / Uncert / Opt Software
High Perf Computing
Training
ESW
Communications
Input Data
Tech Support



Validation Testing
Engine Validation Planning

Vehicle Systems
Probabilistic
Ambient Temp
Distribution

Fan & Compressor
HFB Producibility
Parametric Airfoil
Compressor Aero Design

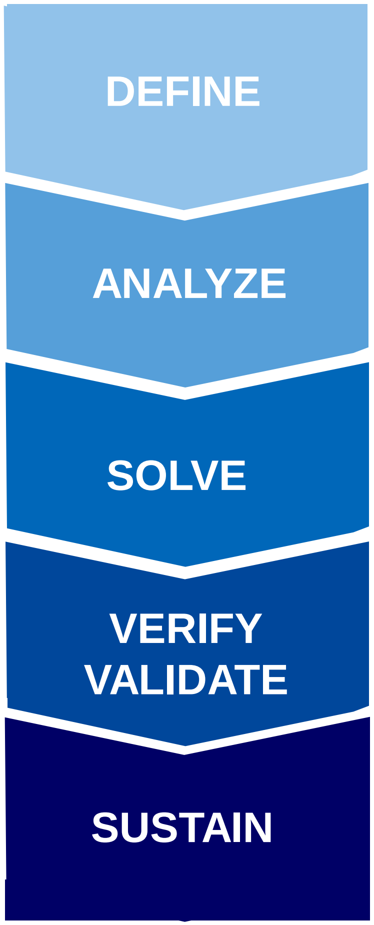
Structures
Probabilistic HCF
Parametric Geometry Simulation Model
Engine Dynamics and Loads

Turbine
Turbine Blade Durability
Turbine Vanes and BOAS Durability
Rotor Thermal Model
Airfoil LCF Lifing

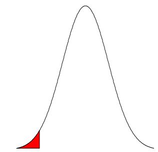
Performance Analysis
Performance Monte Carlo Risk Assessment
Engine Test Confidence, Uncertainty
Uncertainty in Engine System Predictions
Production Test Data Trending and Analysis

Design For Variation

Five Components



DEFINE Customer requirements (probabilistic)



ANALYZE Identify root causes of variation and uncertainty, develop variability/uncertainty model

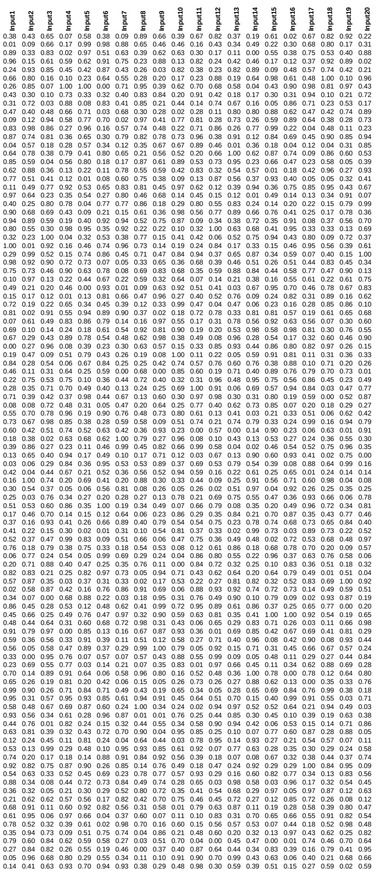
SOLVE Identify ‘optimum’ design that satisfies requirements

VERIFY/VALIDATE Variability/Uncertainty model

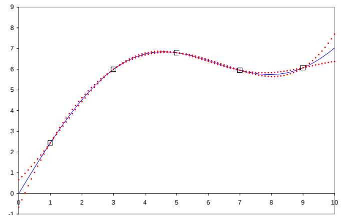
SUSTAIN Stable system of causes of performance variation

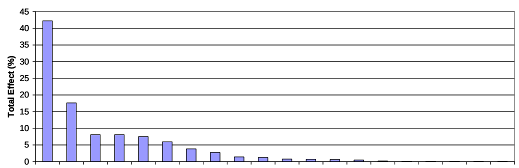
Design For Variation

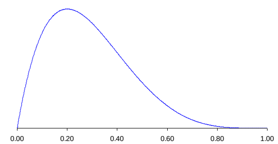
ANALYZE Identify root causes of performance variation and uncertainty and their effects

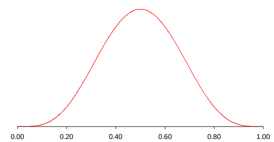


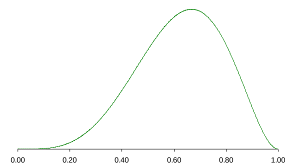
Engineering Model





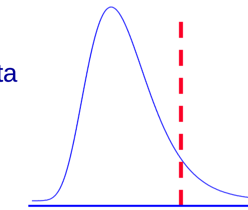






Bayesian Model Calibration

- Parameter uncertainty update
- Bias correction
- Residual variation



Accounting for uncertainty in
- Model input
- Model itself

Design Space Filling Experiment Over Model Input Space

Run Experiment Through Engineering Model

Develop Model Emulator, Sensitivity Analysis

Refine Distributions of Important Model Inputs

Perform Bayesian Model Calibration

Run Real World Uncertainty Analysis

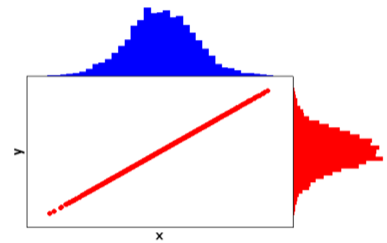
Design For Variation

SOLVE Identify optimum design that satisfies requirements

- Performance characteristic y = $f(x_1, x_2, \ldots, x_p)$ depends on p inputs

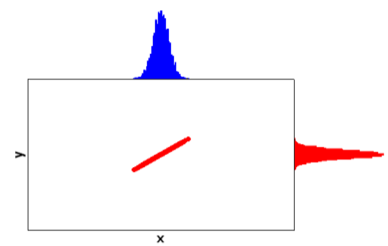
- The variance of y can be approximated by

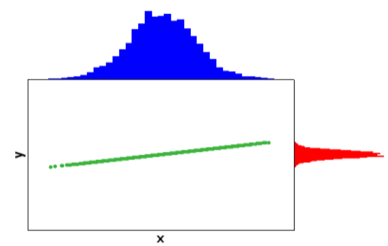
$$\sigma_y^2 \cong \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \cdots + \left(\frac{\partial f}{\partial x_p}\right)^2 \sigma_{x_p}^2$$



- We can reduce σ_y^2 by

1. Reducing $\sigma_{x_i}^2$: the variance in the inputs $x_1, x_2, \ldots, x_p$
2. Reducing $\dfrac{\partial f}{\partial x_i}$: the sensitivity of y to variation in $x_1, x_2, \ldots, x_p$

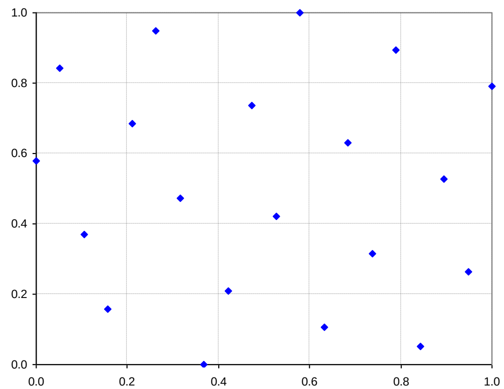




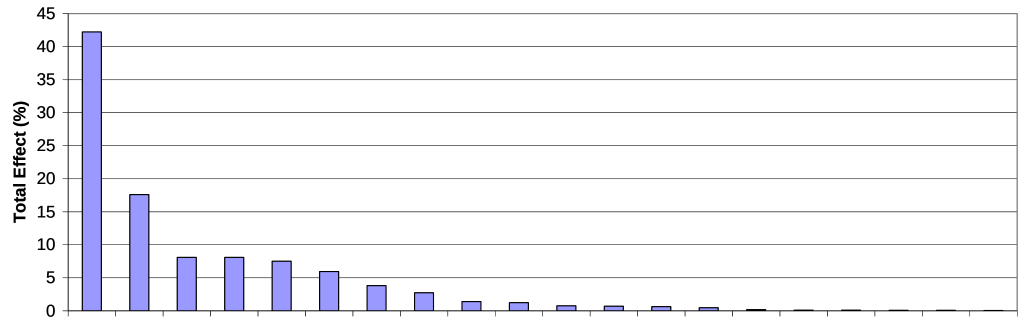
Design For Variation

ANALYZE : Key Technologies

1. Latin Hypercube Experimental Designs



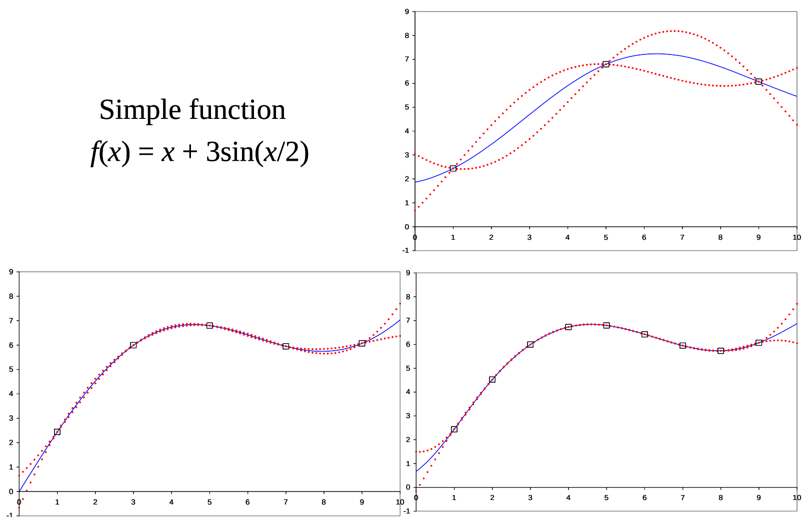
3. Variance-Based Sensitivity Analysis



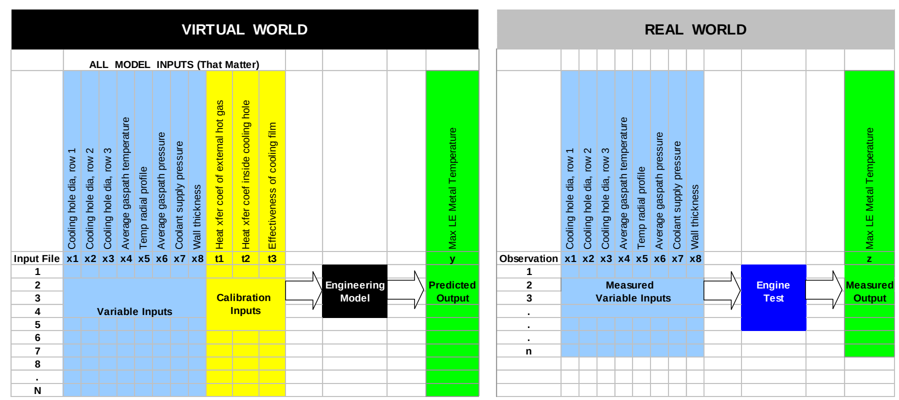
← Model Inputs →

2. Gaussian Process Emulators

Simple function
$f(x) = x + 3\sin(x/2)$



4. Kennedy and O'Hagan Bayesian Model Calibration



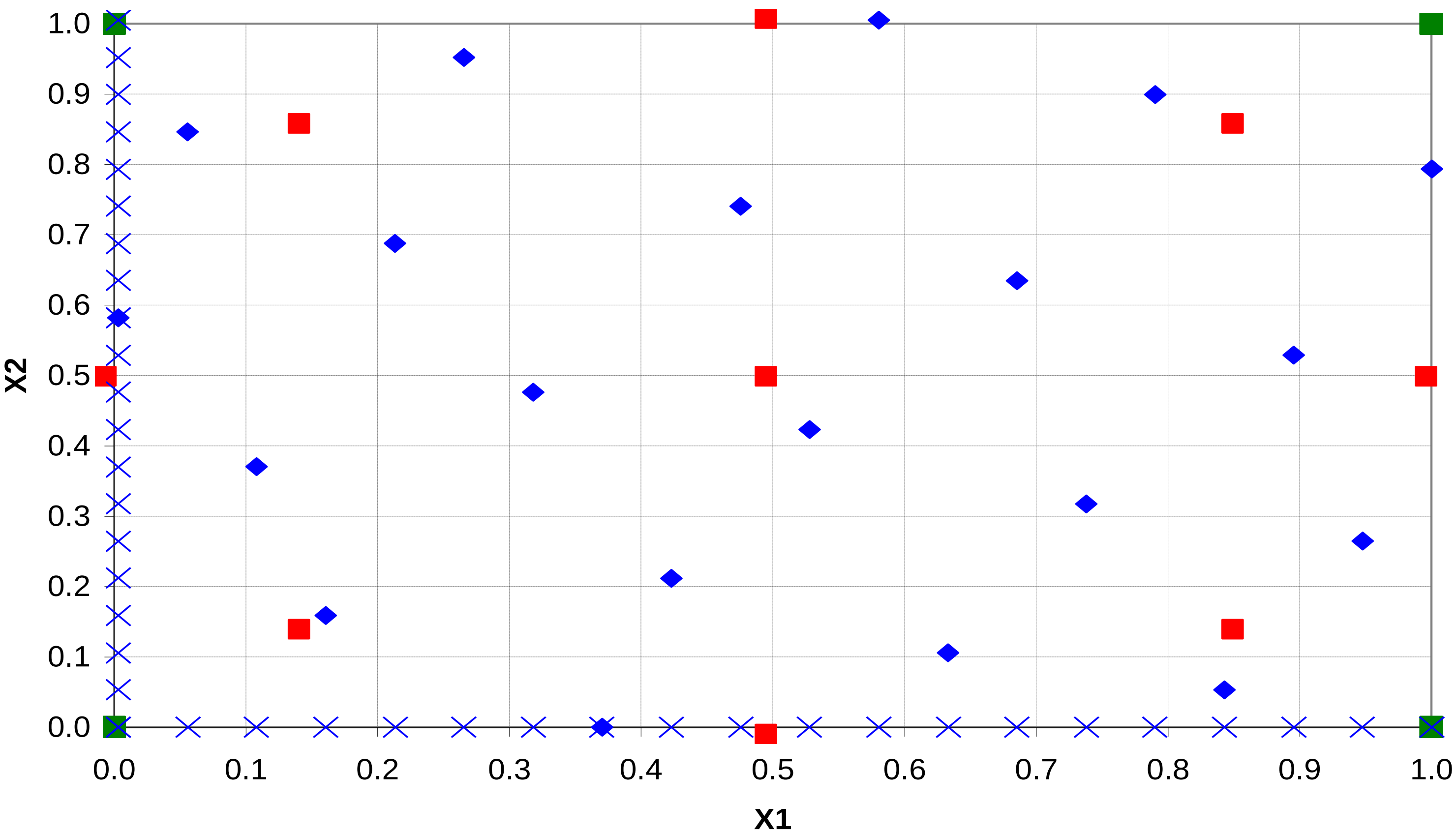
Predictions

Kennedy and O'Hagan model/methods
- Calibrate engineering model
- Quantify model uncertainty
  - ✓ Parameter, bias, residual

Data

Design For Variation

1. Latin Hypercube Experimental Designs



2-level factorial designs assume linearity and focus on vertices of design space

2nd order response surface designs like the **CCD** tend toward corners and edges but improve on factorial designs

Latin Hypercube samples are space-filling and guarantee uniform distribution over margins (see X's in diagram)

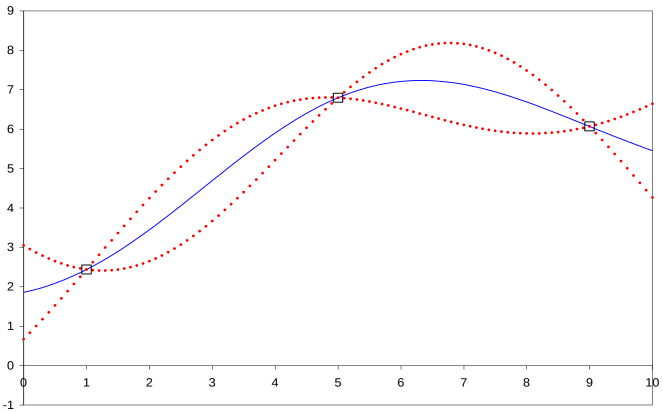
Design For Variation

2. Gaussian Process Emulators

- Thousands of model runs typically required for uncertainty analysis
  - Calibration
  - Propagation of scenario uncertainty
  - Sensitivity analysis
- Not practical for computationally expensive codes
- Gaussian process models as ‘emulators’
  - Approximate model y=f(x)
  - Provide probability distribution quantifying uncertainty at new design points

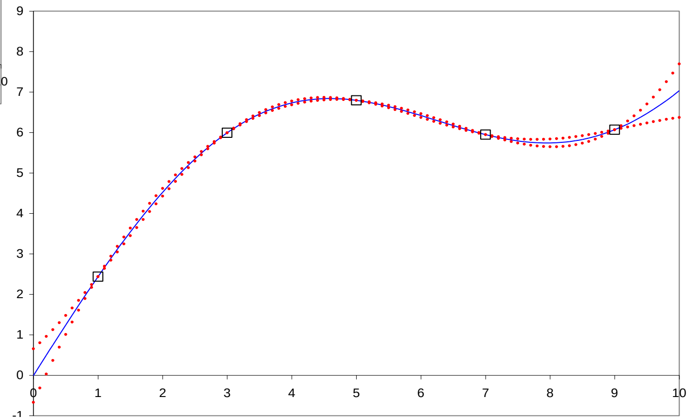
Design For Variation

2. Gaussian Process Emulators



Simple function

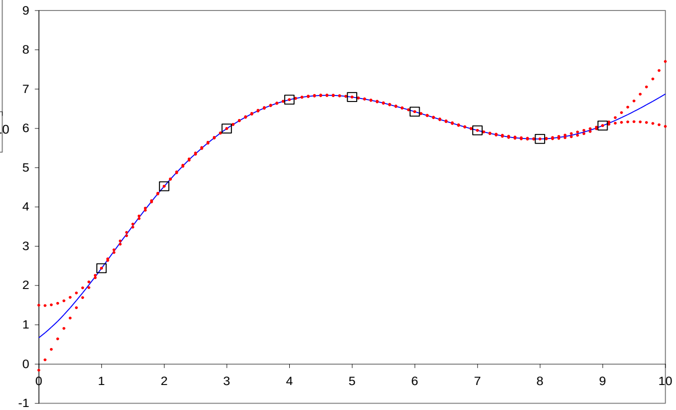
$$f(x) = x + 3\sin(x/2)$$



<u>Observations</u>

1. Zero uncertainty at training data
2. Uncertainty increases with distance from training data
3. Uncertainty decreases as training data are added
4. Emulator shape changes as it "learns" from training data

These are desirable properties of any emulator



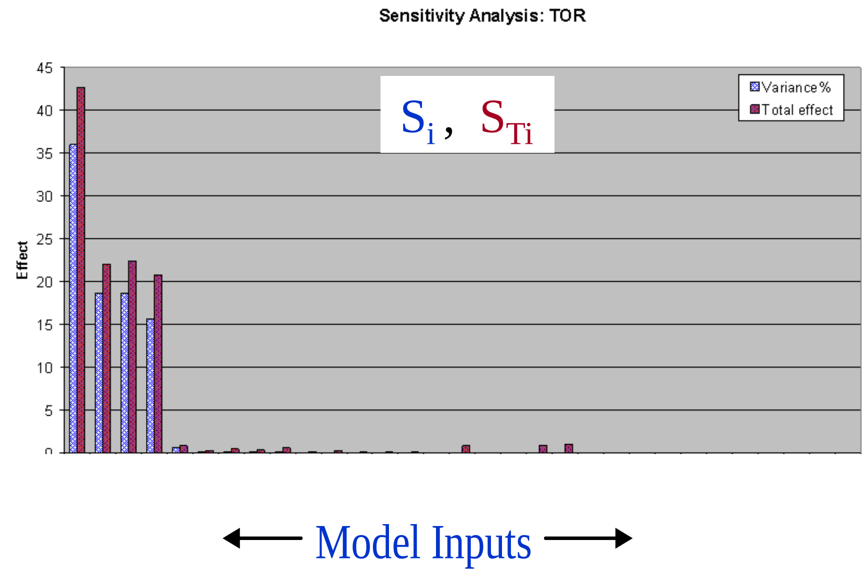
Ref: O'Hagan, A. (2006). Bayesian analysis of computer code outputs: a tutorial. Reliability Engineering and System Safety, 91, 1290-1300.

Design For Variation

3. Variance Based Sensitivity Analysis

How much of the variance of model output is due to each input?

1. $S_i = Var[E(Y|X_i)]/Var(Y)$
   - % Due to Main Effect of X_i
2. $S_{Ti} = E[Var(Y|\boldsymbol{X_{-i}})]/Var(Y)$
   - % Due to Total Effect of X_i
   - Main Effects + All Interaction Effects involving X_i , if X_i independent



Design For Variation

4. Kennedy and O'Hagan Bayesian Model Calibration

What is an Uncertainty Analysis?

- An uncertainty analysis augments a single point prediction with a probability distribution that accounts for
  - Variability or uncertainty in model input
  - Capability of model
- How variable or uncertain is your model input
  - Uncertainty due to random variation in model inputs
- How capable is your model
  - Uncertainty due to lack of agreement between model predictions and physical measurements (the real world response)
  - Model validation

4. Kennedy and O'Hagan Bayesian Model Calibration

Uncertainty

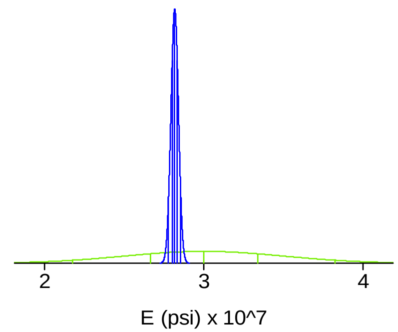
- Some sources of uncertainty in the state of a physical system associated with a deterministic prediction

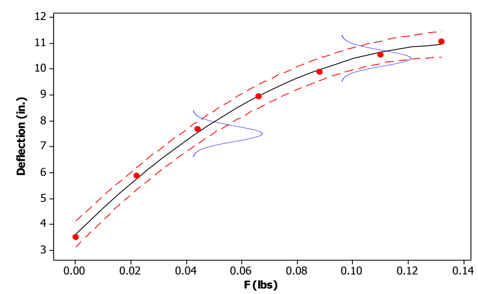
  a. Scenario uncertainty - Uncertainty about some future measurable values of model inputs, e.g., what missions *will* be flown, what hole sizes *will* result from the laser drilling process

  b. Parameter uncertainty - Uncertainty about best values of model parameters (e.g. heat transfer coefficients, Young's modulus, compressor efficiency) or uncertain inputs (e.g. boundary conditions)

  c. Model structure uncertainty - Uncertainty about the difference between the mean of the real world process being modelled and the model prediction using the best possible parameter values. Sometimes referred to as model inadequacy, model discrepancy, or model bias.

  d. Residual variation - Variation in real world outcomes at a given (known) scenario, due to variation in factors that are outside the model or measurement error

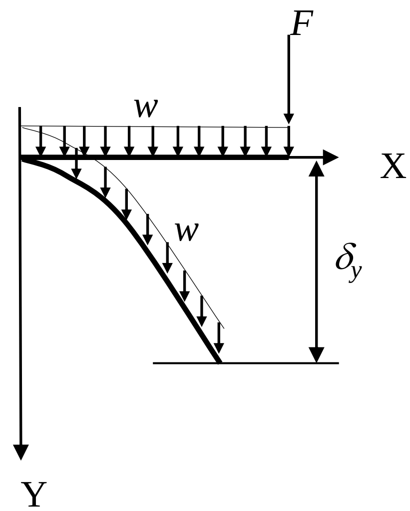


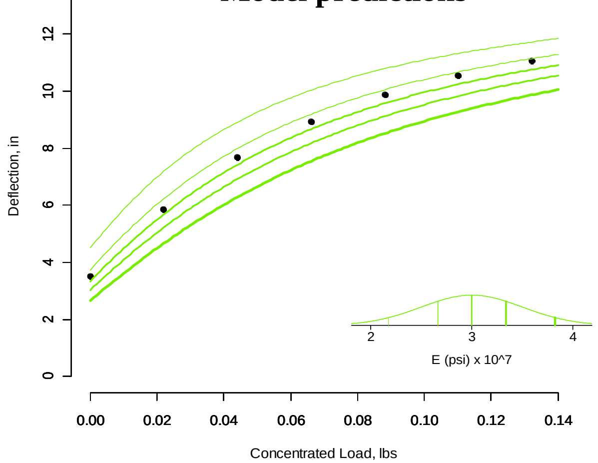


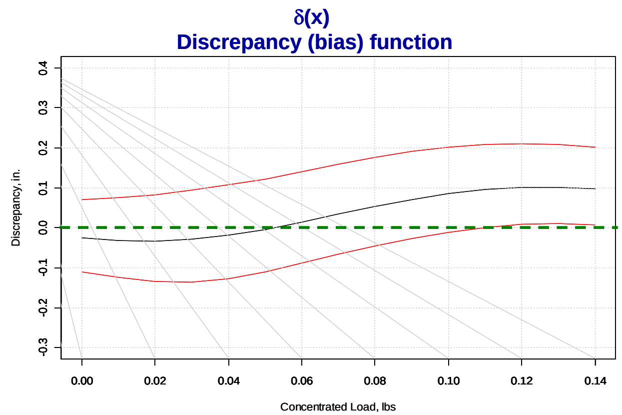
4. Kennedy and O'Hagan Bayesian Model Calibration

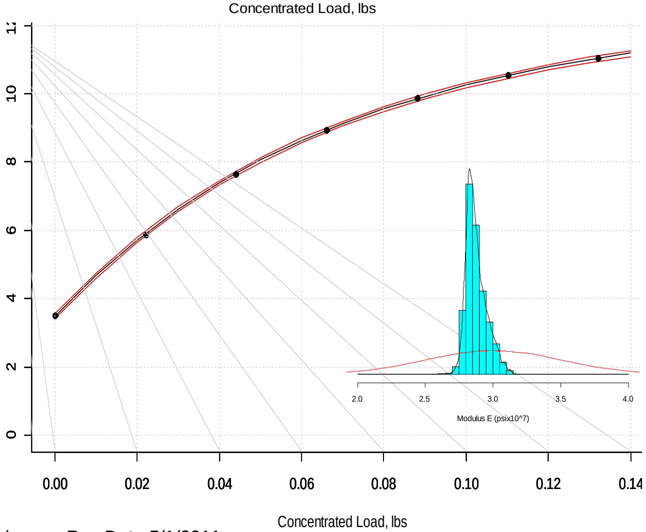
A statistical framework for combining experimental data with model predictions to provide best estimates and uncertainty for

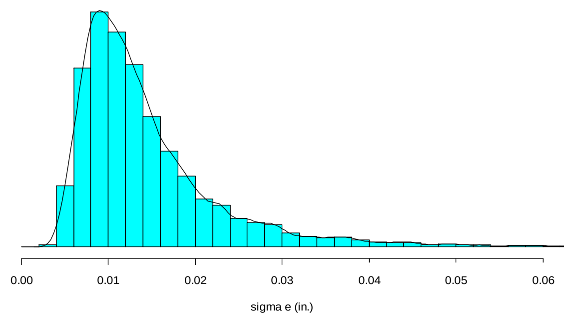
- **Model calibration parameters**
- **Systematic discrepancies between model and data**
- **Standard deviation of random discrepancies between model and data**
- **Model predictions**











Nomenclature

x	**Model variable (measurable) inputs (exper. conditions)**
θ*	**Model calibration inputs**
θ	**Best value of model calibration inputs**
ζ(x)	**True average system response given inputs x**
η(x,θ*)	**Model prediction for inputs x and θ***
y(x)	**Experimental observation for inputs x**
δ(x)	**discrepancy (bias) between ζ(x) and η(x,θ)**
e(x)	**random observation error of the experimental data**

Data few and noisy but unbiased	$y(x) = \zeta(x) + e(x)$
Model smooth but biased	$\zeta(x) = \eta(x,\theta) + \delta(x)$
Combine to get best from both	$y(x) = \eta(x,\theta) + \delta(x) + e(x)$

Design For Variation

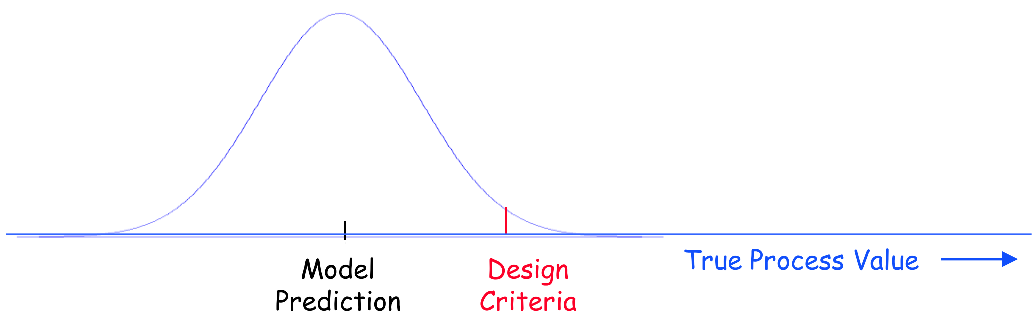
Systematic Process for Designing for and Managing Uncertainty and Variability

- Establish probabilistic design requirements
- Emulate, calibrate engineering models
- Solve for design that meets probabilistic requirements
  - Look for opportunities for making design less sensitive to variation
- Validate and sustain model
- Write Engineering Standard Work, develop local training

Design For Variation

- Goal: quantify, understand, and control the risk of not meeting design criteria or exceeding thresholds

- "The revolutionary idea that defines the boundary between modern times and the past is the mastery of risk: the notion that the future is more than a whim of the gods and that men and women are not passive before nature."
  - *Peter Bernstein, "Against the Gods: The remarkable story of risk"*



Bayesian Model Calibration

Challenges

- Establishment of 'Gold Standard' numerical methods
- Commercial software availability
- Parametric geometry
- Optimal [model:experimental] DOE for model validation
- Computational issues (e.g. matrix inversion $O(n^3)$)
- Large transient models
- Calibration data outside operational range
- What if only sub-models can be calibrated?
- Discrepancy root cause investigation structure
  - Original research assumed measurement process free of bias
  - Sometimes instrumentation technology can rival model technology
- Best approach to confounding issues
- Lack of textbooks, engineering methods and applications papers

Design For Variation

Bibliography

Websites

1. http://www.mucm.ac.uk/Pages/Resources/Resources_ReadingList.htm, An excellent annotated bibliography
2. http://mucm.group.shef.ac.uk/Pages/Dissemination/Dissemination_Presentations.html, presentations and talks at MUCM
3. http://www.stat.duke.edu/~fei/samsi/index.html
4. http://legacy.samsi.info/programs/2006compmodprogram.shtml

Uncertainty, Bayesian Statistics

5. Lindley, D.V. (2006). *Understanding Uncertainty*. New York: John Wiley and Sons.
6. O'Hagan, A. (2004). Dicing with the unknown. *Significance*, Volume 1, Issue 3, pages 132–133
7. Bernardo, J. M. (2003). Bayesian Statistics. *Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics*, (R. Viertl, ed). Oxford, UK: UNESCO (www.eolss.net)
8. Lindley, D. V. (2000). The Philosophy of Statistics. *Journal of the Royal Statistical Society*: Series D (The Statistician), Vol. 49, No. 3, pp. 293–337
9. Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis,Second Edition*. New York: Chapman and Hall.
10. O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite,P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). *Uncertain Judgements: Eliciting Expert Probabilities*. John Wiley and Sons,Chichester.

Bayesian Model Calibration, Methods

11. O'Hagan, A. (2006). Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety* 91, 1290–1300.
12. Kennedy, M. C. and OHagan, A. (2001). Bayesian calibration of computer models (with discussion).*J.R.Statist.Soc.B*, 63, 425-464.
13. M. C. Kennedy and A. O'Hagan 2001. Supplementary details on Bayesian calibration of computer models, *Internal report*, University of Sheffield.
14. Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., and Tu, J. (2007b), A Framework for Validation of Computer Models, *Technometrics*, 49, 138–154.
15. Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J., and Walsh, D. (2007a), Computer Model Validation with Functional Output. *Annals of Statistics*, 35, 1874–1906.
16. Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), Combining Field Data and Computer Simulation for Calibration and Prediction. *SIAM Journal on Scientific Computing*, 26, 448–466.
17. Williams, B., Higdon, D., Gattiker, J., Moore, L., McKay, M., and Keller-McNulty, S. (2006), Combining Experimental Data and Computer Simulations, With an Application to Flyer Plate Experiments, *Bayesian Analysis* 1, Number 4, pp. 765-792
18. Higdon, D., Gattiker, J., Williams, B., Rightly, M. (2008), Computer Model Calibration Using High-Dimensional Output, *Journal of the American Statistical Association*, 103, 570-583.

Design For Variation

Bibliography

Gaussian Process Emulation, Experimental Design for Emulation

19. Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.
20. Fang, K., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Boca Raton, FL: Chapman and Hall/CRC
21. Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering Design via Surrogate Modelling: A Practical Guide*, New York: Wiley.
22. Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA: The MIT Press.
23. Bastos, L. and O'Hagan, A. (2009). Diagnostics for Gaussian Process Emulators, *Technometrics* 51, 425-438.
24. Oakley, J.E. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society* B 66, 751-769.
25. Gramancy, R. B., and Lee, H. K. H. (2008), Bayesian Treed Gaussian Process Models With an Application to Computer Modeling, *Journal of the American Statistical Association*, 103, 1119–1130.
26. Sacks, J., Schiller, S. B., and Welch, W. J. (1989a), Designs for Computer Experiments, *Technometrics*, 31, 41–47.
27. Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M.D.(1992), Screening, Predicting, and Computer Experiments, *Technometrics*, 34, 15–25.
28. Tang, B. (1993), Orthogonal Array-Based Latin Hypercubes, *Journal of the American Statistical Association*, 88, 1392–1397.
29. Fang, K. T., Lin, D. K. J., Winker, P., and Zhang, Y. (2000), Uniform Design: Theory and Application, *Technometrics*, 42, 237–248.
30. Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), Minimax and Maximin Distance Designs, *Journal of Statistical Planning and Inference*, 26, 131–148.
31. Loeppky, J., Sacks, J., Welch, W (2009). Choosing the Sample Size of a Computer Experiment: A Practical Guide, *Technometrics.* November 1, 2009, 51(4): 366-376

Computational Methods (Modular Bayes, MCMC)

32. Liu, F., Bayarri, M., and Berger, J. (2009). Modularization in Bayesian Analysis, with Emphasis on Analysis of Computer Models, *Bayesian Statistics* 4, Number 1, pp. 119-150.
33. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, 21, 1087–1091.
34. Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo*. Boca Raton, FL: Chapman and Hall
35. Gilks W.R., Richardson S. and Spiegelhalter D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC.
36. Albert, J. (2009). *Bayesian Computation with R*, New York: Springer
37. R: http://www.r-project.org/

Pratt & Whitney
A United Technologies Company

Design For Variation

Bibliography

Applications

38. Kennedy, M.C., Anderson, C.W., Conti, S. and O'Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. *Reliability Engineering and System Safety* 91, 1301–1309.

39. Aslett, R., Buck, R. J., Duvall, S. G., Sacks, J., and Welch, W. J. (1998), Circuit Optimization via Sequential Computer Experiments: Design of an Ouput Buffer, *Applied Statistics*, 47, 31–48.

40. Bayarri, M. J., Berger, J. O., Kennedy, M. C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Lin, C. H., and Tu, J. (2005), Bayesian Validation of a Computer Model for Vehicle Crashworthiness, *Tech. Rep. 163, National Institute of Statistical Sciences*, Research Triangle Park, NC, http://www.niss.org/technicalreports/tr163.pdf.

41. Chang, P. B., Williams, B. J., Bhalla, K. S. B., Belknap, T. W., Santner, T. J., Notz, W. I., and Bartel, D. L. (2001), Design and Analysis of Robust Total Joint Replacements: Finite Element Model Experiments With Environmental Variables, *Journal of Biomechanical Engineering*, 123, 239–246.

42. Liu, F., Bayarri, M. J., Berger, J. O., Paulo, R., and Sacks, J. (2008), A Bayesian Analysis of the Thermal Challenge Problem, *Computational Methods and Applications in Mechanical Engineering,* to appear, also available as National Institute of Statistical Sciences Technical Report 166, http://www.niss.org/technicalreports/tr166.pdf.

43. Heitmann, K., Higdon, D., Nakhleh, C. and Habib,S. (2006). Cosmic Calibration LAUR-06-2320

44. Anderson, B. and Haller, H. (2011). Combining Computational Fluid Dynamics Analysis Data with Experimentally Measured Data, *Quality Engineering* 23, 46-58.

Pratt & Whitney
A United Technologies Company

Kennedy and O'Hagan Bayesian Model Calibration

Sample of Available Software [Preliminary]

Software	Data Analysis	Bayesian Data Analysis	Space-Filling DOE	Model Emulation*	Sensitivity Analysis	Calibration, single response	Calibration, multiple responses	Optimization	Uncertainty Analysis
GEMSA			X	X	X				X
GEMCAL						X			
GPMSA			X	X	X	X	X		
DAKOTA			X	X	X			X	X
Isight			X					X	X
Matlab (Optimization Toolbox)								X	
Matlab (Statistics Toolbox)	X		X						X
R (BACCO Package)				X	X	X			
R (Base Package)	X							X	X
R (gptk Package)				X					
R (lhs, DiceDesign, DiceKriging Packages)			X						
R (mlegp Package)				X	X				
R (tgp Package)				X	X				
R (Sensitivity Package)					X				
R (LearnBayes package)		X							
WinBUGS		X							
JMP	X		X	X	X				X
SAS	X	X						X	X
Minitab	X								
SimLab (Matlab/R)					X				X

MUCM Toolkit (algorithms): http://mucm.aston.ac.uk/MUCM/MUCMToolkit/index.php?page=MetaHomePage.html

*Various Independent softwares exist for Model Emulation. See http://www.gaussianprocess.org/ for a listing

Note: GPMSA available through Brian Williams at LANL: brianw@lanl.gov

NASA STATISTICAL ENGINEERING SYMPOSIUM

A RELIABILITY-BASED TOOL FOR LIFE LIMIT EXTENSION OF THE SPACE SHUTTLE MAIN ENGINE (SSME)

A SPACE SHUTTLE LESSON LEARNED

Fayssal M. Safie, Ph. D.

NASA R&M Engineering Technical Fellow

May 3-5, 2011

Outline

- Introduction
- The Need for the Tool
- The Mathematical Bases
- The Tool
  - Assumptions
  - The process
- The Application
- Concluding Remarks

Introduction – Related Material

- “A Criterion for Establishing Life Limits”, 1990, by Gill Skopp and Al Porter.
- "A Statistical Approach for Risk Management of Space Shuttle Main Engine Components“, 1991 Probabilistic Safety Assessment and Management Conference, Beverly Hills, CA, by Fayssal M. Safie.
- “Lower Bound on Reliability for Weibull When Shape Parameter is not Estimated Accurately”, 1991, by Zhoa Huang and Al Porter.
- “Weibull Analysis Handbook”, 1983, by R. Abernathy, C. Medlin, and G. Reinman.

Introduction

- This work was done as part of the National Aeronautics and Space Administration (NASA) effort to introduce the use of Statistical/probabilistic models in managing the risk for critical Space Shuttle hardware.
- The result was a development of a statistically-based risk management tool to consistently and effectively extend the life limit of the Space Shuttle Main Engine (SSME) hardware based on the operational history combined with other engineering information.
- The purpose of the tool was to provide a standardized approach to disposition structural life limitations.
- The tool is called the Single Flight Reliability (SFR) criterion.

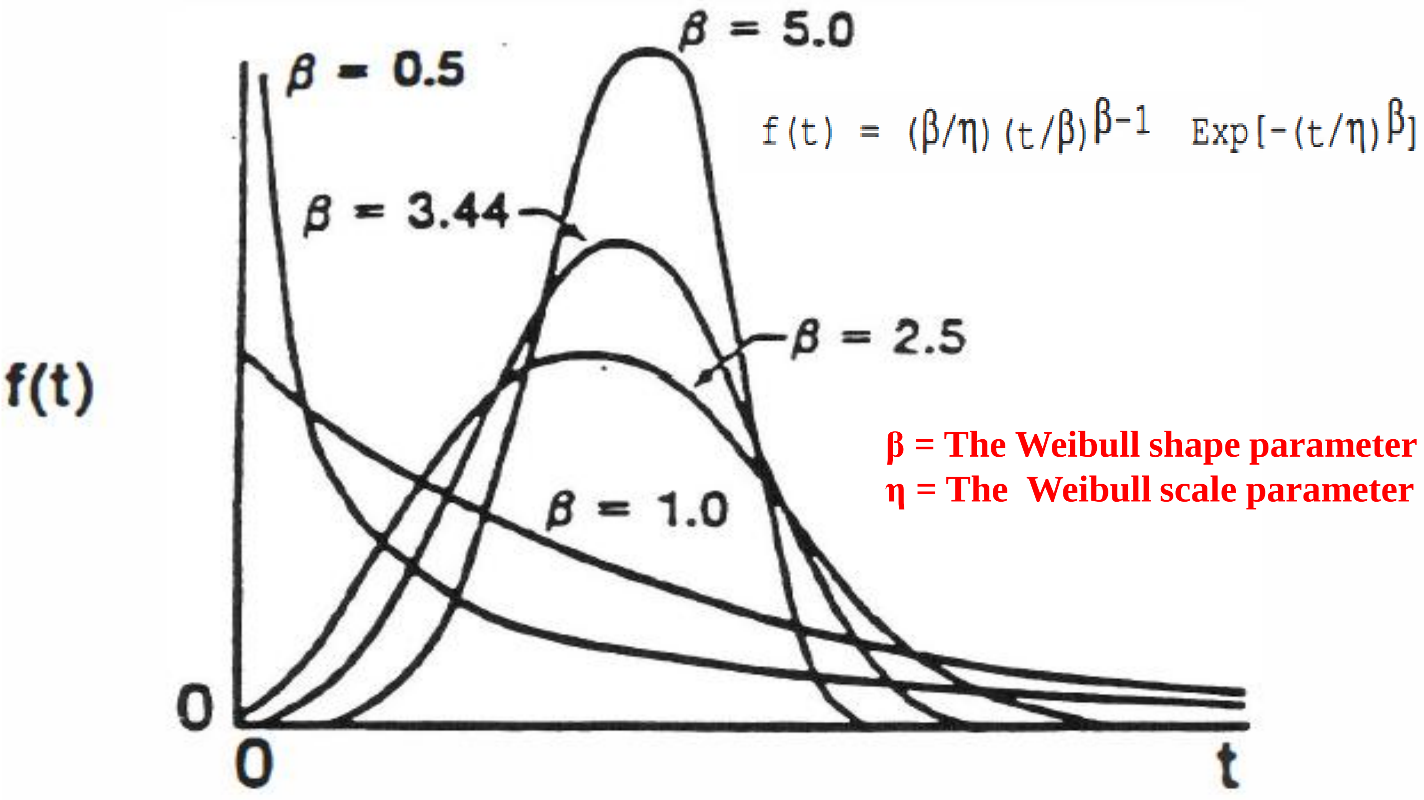
The Need for The Tool

NASA SAFETY CENTER

The SSME Standard Flight Deviation Approval Request (DAR) Criteria

PARAMETER	I	II SFR	III	IV
DAR BASIS	ANALYSIS	OPERATING HISTORY	ANALYSIS	OPERATING HISTORY
MATERIAL PROPERTIES	PREDICTED MINIMUM /EXPECTED MINIMUM	UNIQUE CONDITIONS	EXPECTED MINIMUM	UNIQUE CONDITIONS
OPERATING STRESSES	PREDICTED LOADS MEASURED LOADS/ STRESSES EXTRAPOLATED LOADS/STRESSES	UNIQUE CONDITIONS	MEASURED LOADS/ STRESSES • CORRELATED TO 99/95	UNIQUE CONDITIONS
LIFE LIMITATIONS	BASED ON ANALYSIS	25% FLEET LEADER OR 2 STATIS. JUSTIFIED ≤ 50% F/L 6 UNITS ≥ LIMIT	BASED ON ANALYSIS ≤ 50% F/L	≤50% FLT. LEADER/ FAILED UNIT 6 UNITS ≥ LIMIT
PERIODIC INSPECTION	NONE	NONE	25% F/L EXPOSURE	25% LIFE INTERVAL
LIFE FACTOR	HCF: 10, PRED. MIN. 4, EXP. MIN. LCF: 4	4 2 OR, 0.995/0.90 (RELIABILITY/ CONFIDENCE)	HCF: 2 LCF: 4	2

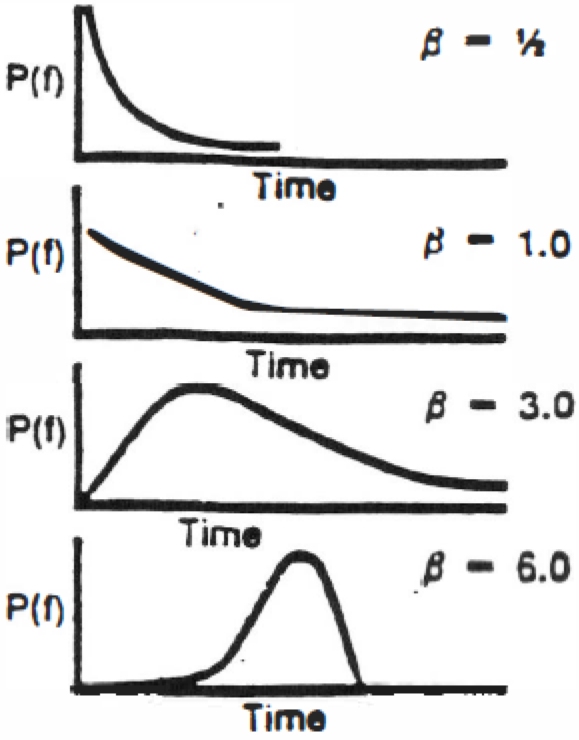
The Mathematical Basis
The Weibull Probability Density Function



β = The Weibull shape parameter
η = The Weibull scale parameter

The Mathematical Basis
The Significance of the Weibull Shape Parameter

1. Infant mortality
   - Inadequate burn-in, green run
   - Misassembly
   - Some quality problems
2. Random failures
   - Independent of time
   - Maintenance errors
   - Electronics
   - Mixtures of problems
3. Early wearout
   - Surprise!
   - Low cycle fatigue
4. Old age wearout (rapid)
   - Bearings
   - Corrosion



The Mathematical Bases – The Equations

The Weibull probability density function:

$$f(t) = (\beta/\eta)(t/\beta)^{\beta-1} \; \mathrm{Exp}[-(t/\eta)^{\beta}]$$ **(two parameters)**

The Weibull reliability and failure functions:

$$R(t) = \mathrm{Exp}[-(t/\eta)^{\beta}] \quad \text{(two-parameters)}$$

$$F(t) = 1 - \mathrm{Exp}[-(t/\eta)^{\beta}]$$

β = The Weibull shape parameter
η = The Weibull scale parameter
t = The total time
m = the single mission time
(1-α)*100 = The confidence level

The Weibayes

$$\eta \geq \left(\sum_{i=1}^{n}(t_i^{\beta})/-\ln \alpha\right)^{1/\beta}$$

90% Confidence

$$\eta \geq \left(\sum_{i=1}^{n}(t_i^{\beta})/2.3\right)^{1/\beta}$$

The Weibull conditional probability function:

$$P(T \geq t \mid T > t-m) = \mathrm{Exp}[-(t/\eta)^{\beta}]/\mathrm{Exp}[-((t-m)/\eta)^{\beta}]$$

The Tool - Assumptions

- Infant mortality situations are excluded .
- For a specific SSME component, all units have the same basic configuration and geometry, and all are tested in the same environment.
- Only SSME components with extensive fleet hot fire experience with no failure history are considered.

The Tool – The Process

- The SFR Criterion uses a statistical approach to derive a life limit for a given component subject to a specified reliability and confidence level requirement.
- The statistical approach developed is based on Weibull time-to-failure distribution.
- Since the SFR Criterion applies only to components with no failures and the shape parameter of the Weibull distribution varies for different components, the Weibayes and a conditional Weibull reliability functions were used in combination with an optimization technique to derive a minimum life limit.

The Tool - The Process (continued)

Calculating The Minimum Life:

1) Assume a value of (β) of approximately one.
2) For the operational history of the item under consideration, estimate (η) at the 90% confidence level using:

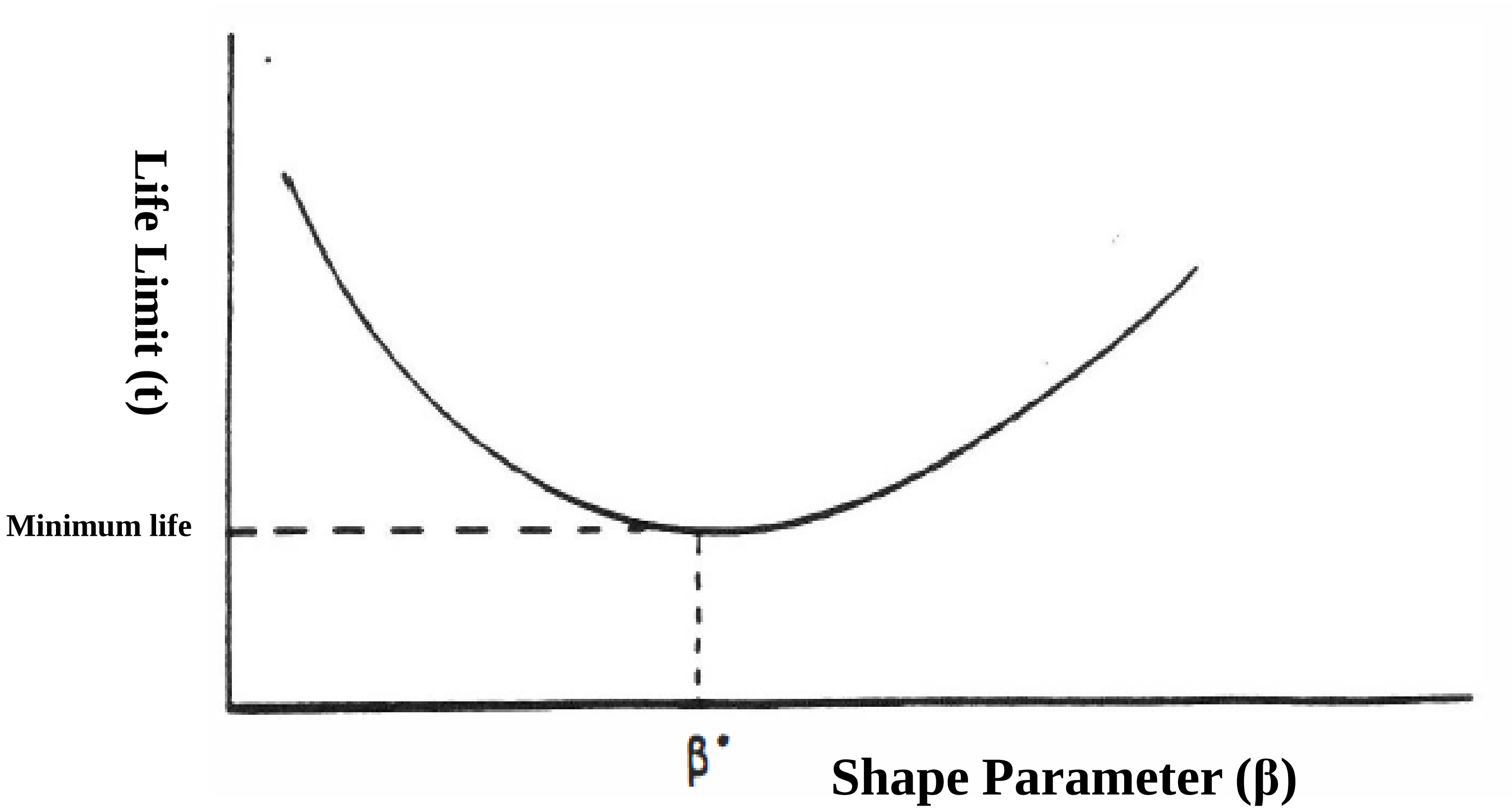
$$\eta \geq \left(\sum_{i=1}^{n}(t_i^{\beta})/2.3\right)^{1/\beta}$$

3) Use the β and η in steps 1 and 2, and the specified single flight reliability (i.e., 0.995) to determine the value of t using:

$$P(T \geq t \mid T > t-m) = \mathrm{Exp}[-(t/\eta)^{\beta}]/\mathrm{Exp}[-((t-m)/\eta)^{\beta}]$$

4) Starting from the second iteration, check if the value of t obtained in step 3 is higher than the value of t obtained from the previous iteration. If so, go to step 6 .
5) Increment the value of β and go to step 2.
6) The value of t is the minimum.

The Mathematical Bases – The Minimum Life



The Tool - The Process (Continued)

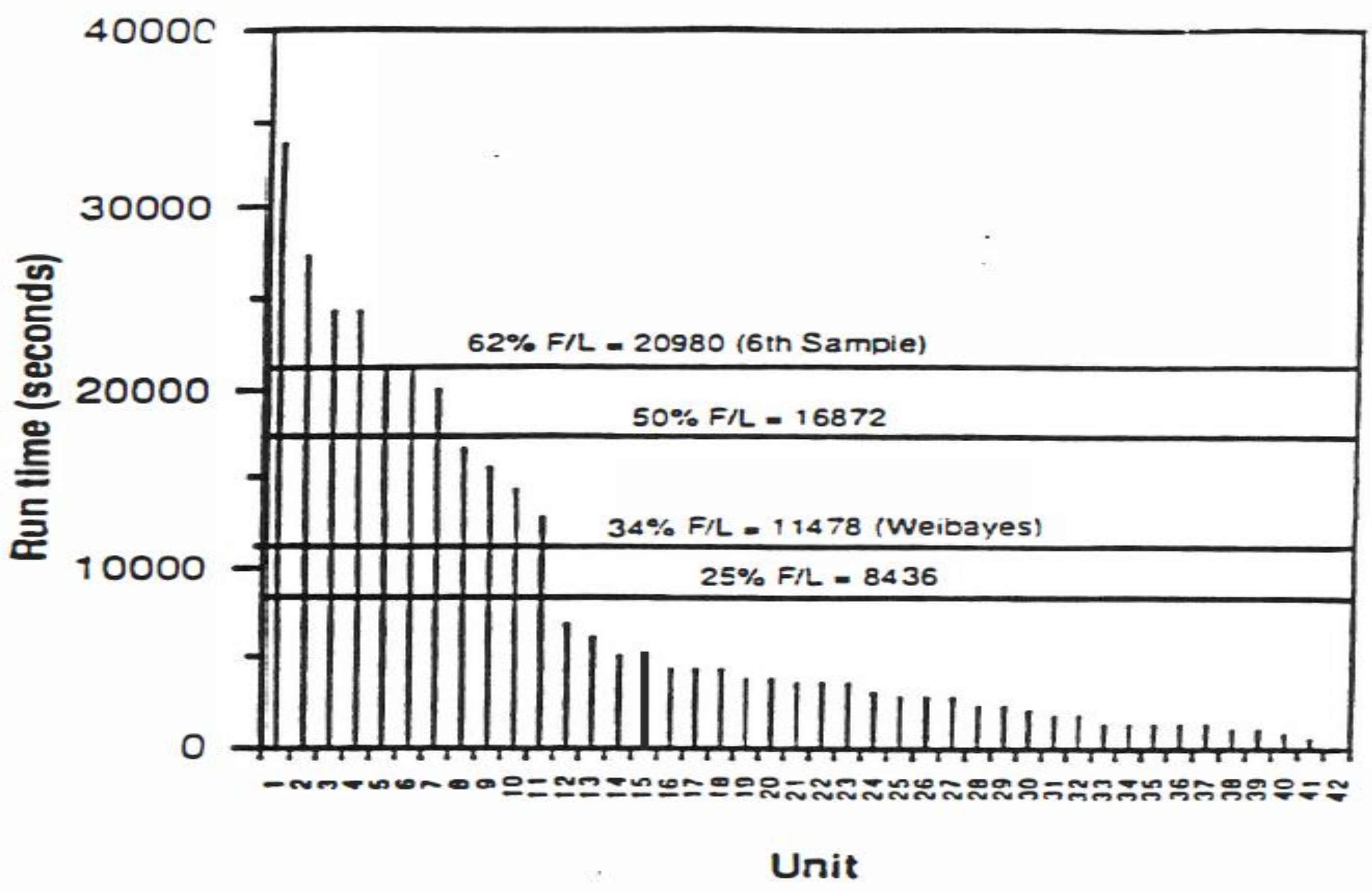
- **The minimum life limit derived is then checked to make sure that it does not exceed 50% of the operating time of the fleet leading unit, or the minimum operating time of the six leading units.**
- **If the life limit derived is less than 25% of the fleet leading unit, the life limit is increased to 25%.**
- **The life limit derived has a lower bound of 25% of the fleet leading unit and an upper bound defined by the lesser of 50% of the fleet leading unit or the lowest of the six leading units.**

The Application - The SSME Fuel Bleed Duct

- Data on 42 SME fuel bleed duct units with zero failures are used here to illustrate the application of the SFR tool.
- Using this data, for a 0.995 single flight reliability and 90% confidence level requirement, the minimum total time, t, derived is 11,478 seconds.
- This value of t represents approximately 34% of the operational experience of the fleet leading unit of 33,744 seconds.
- The 34% is higher than the lower bound of 25% (8,437 seconds) and lower than the upper bound of 50% (16,87 2 seconds) and the minimum of the six leading units (62% of the fleet leader) .
- Therefore, the life limit is 11,478 seconds.

The Application - The SSME Fuel Bleed Duct

The SSME Fuel Bleed Duct



Concluding Remarks

- The statistical tool presented was implemented as part of the Space Shuttle Program requirement.
- The tool has been effectively used by the Shuttle Program since the early 1990’s.
- This is a good example of how Statistical Engineering has helped the SSME program to reduce cost, increase availability, and maintain high level of reliability of critical hardware.

Applications of Bayesian Statistical Analyses in Determining the Number of Demonstration Tests to Conduct and in Monitoring Reliability Growth

Dr. William E. Vesely
Office of Safety and Mission Assurance
NASA Headquarters
Washington, DC 20546

The Basic Problem Addressed

- A new system such as a new spacecraft is to be evaluated for its reliability*
- Part of the evaluation involves determining the number of tests to perform before acceptance
- The evaluation also involves dynamically tracking the reliability evolution of the system with test and operation
- To optimize resources, the evaluations need to utilize all available information
- Uncertainties also need to be treated and be quantified

*Safety is treated as part of reliability here

Basic Concepts: Design Reliability and Demonstration Tests

- Design reliability is the probability that a new system has no inherent failure-causing faults
- Demonstration tests are conducted to detect any such inherent failure-causing faults
- Demonstration tests can be partial tests or can be test flights
- A major issue: How tests are needed to demonstrate a given reliability to a given certainty?

Factors Determining Number of Tests to Conduct

- From reliability growth principles the required number of tests depends on three major factors:
  - Initial Assurance Level
  - Fault-Detection Effectiveness
  - Corrective Action Effectiveness
- Objective: Develop an approach that incorporates these factors and quantifies the reliability after a given number of tests

Framework of the Bayesian Approach

- The reliability estimate is described by a probability distribution to account for uncertainties
- The distribution gives the mean, median, and uncertainty bounds
- An initial distribution (prior distribution) is constructed to account for initial knowledge
- The distribution is updated from the results of a test using Bayes theorem
- This updating is continued to determine the number of required tests or to track performance

Advantages of the Bayesian Approach

- The Bayesian approach can utilize both quantitative and qualitative information
- Uncertainties are comprehensively quantified
- Assessments are dynamically updated as information is gained from the tests
- The Bayesian approach is standardly used in NASA risk and reliability assessments
- Software exists that can carry out the evaluations in an efficient manner

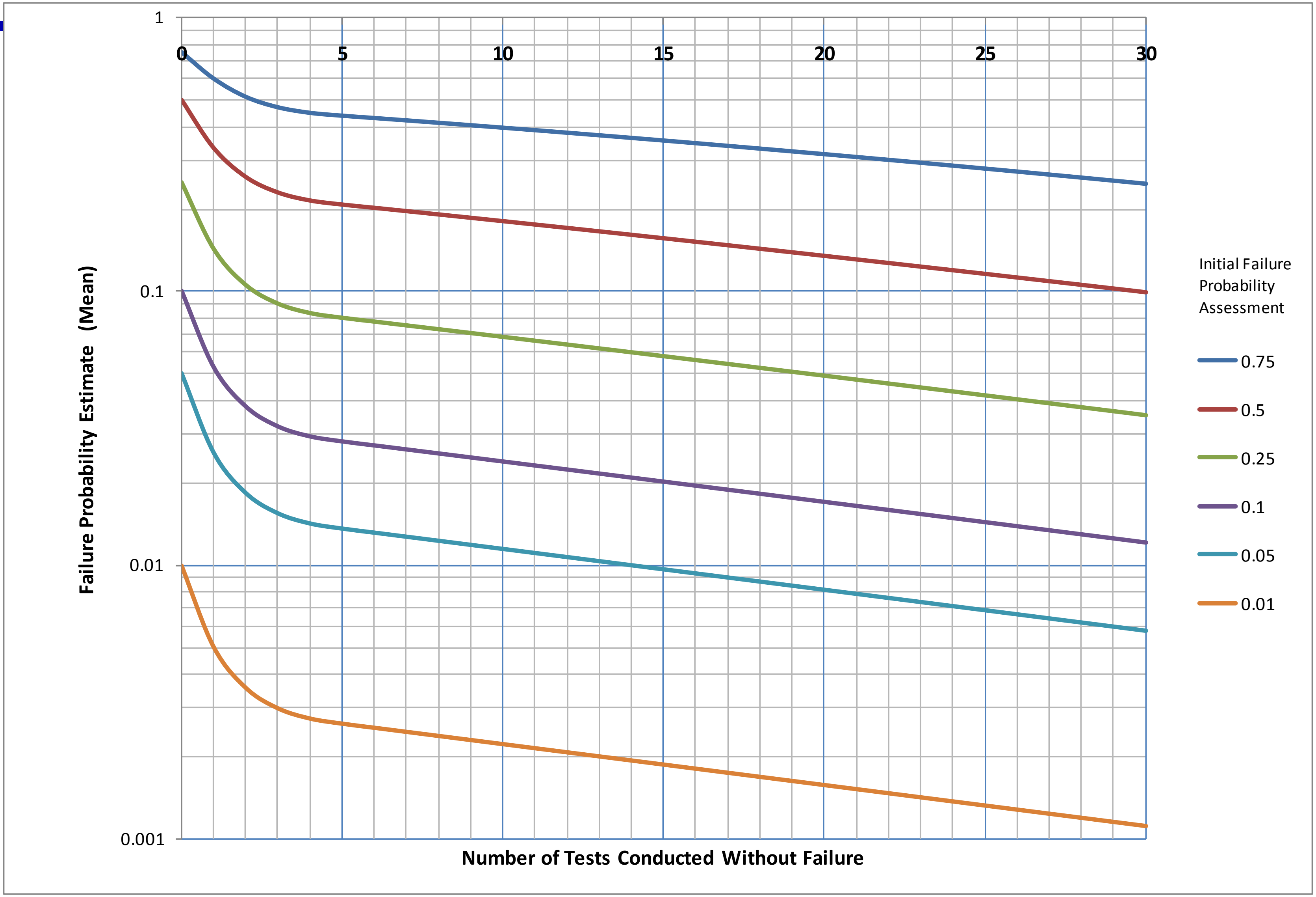
Inputs to the Bayesian Approach

- The *Prior System Reliability Estimate* is determined from the bases for the Initial Assurance Level:
  - Hazard analyses and FMEAs
  - Reliability and Risk analyses
  - Oversights and Reviews
- The *Fault Detection Probability* and the *Fault Correction Probability Estimates* are determined using test and repair information:
  - System specific data
  - Shuttle analyses and data
  - Constellation analyses

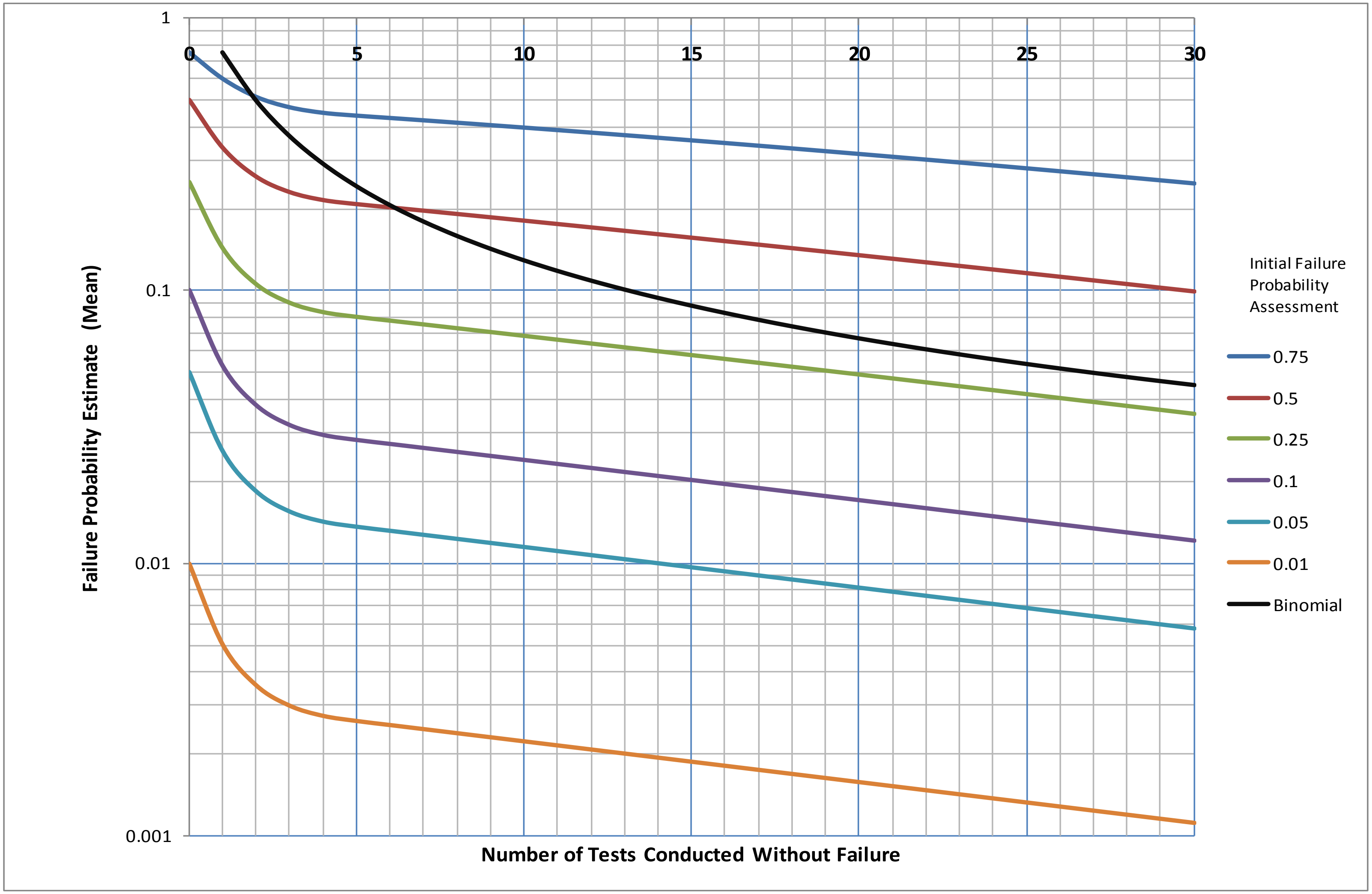
Applications to Determine the Number of Required Failure-Free Tests

- The next slide gives the number of required failure-free tests as a function of the initial assurance level
- The second slide overlays the curve for the binomial estimate which inaccurately treats the tests as throw-away tests
- The third slide quantifies the uncertainty and shows how it decreases with tests even if the initial information is uncertain
- These slides show the decision-making information provided using the Bayesian approach
- These results can be extended to cover the cases where failures or faults occur during the testing

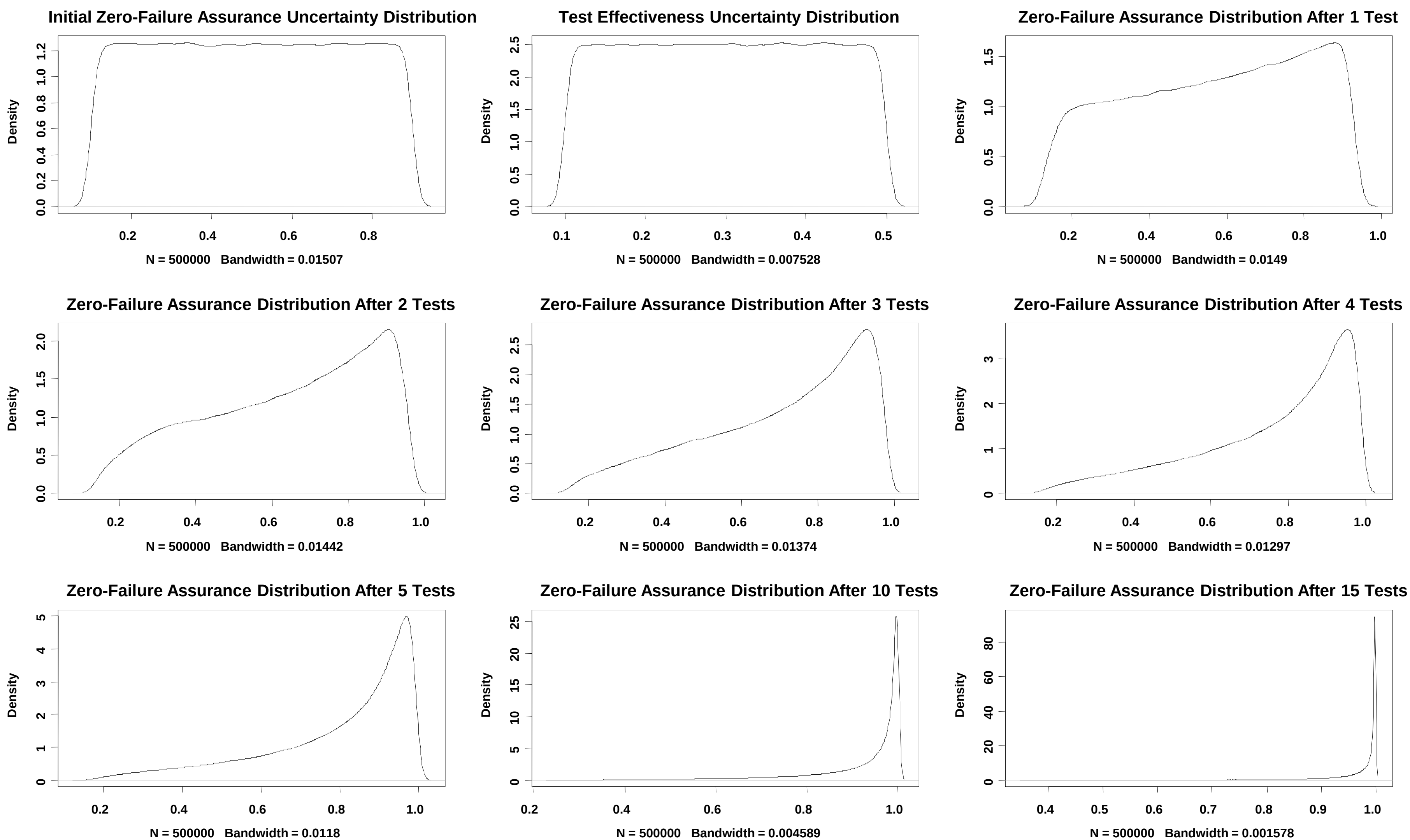
Failure Probability Estimates Versus Number of Failure Free Tests and Initial Failure Probability Estimate



Failure Probability Estimates Compared to the Binomial Sampling Estimate Which Models the System Tests as Throw-Away Sample Tests

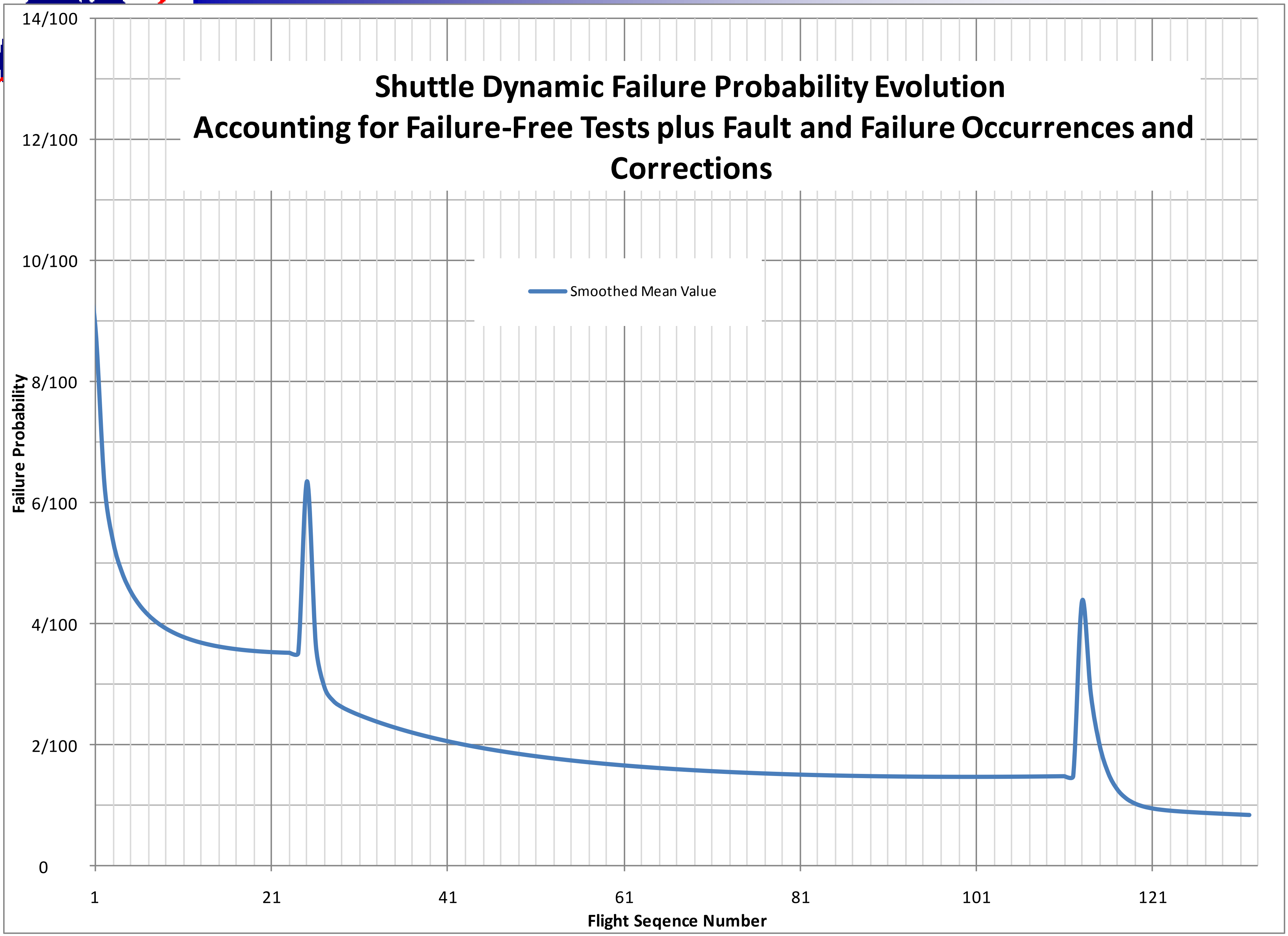


Uncertainty Distribution for the Reliability Estimate After Given Numbers of Tests



Application to Track Reliability Growth

- The next slide shows the application to track the reliability growth of the Space Shuttle
- The application updates the estimate for each next flight based on flight history information
- A Kalman filter is basically used on a transformed scale with Bayesian updating
- Both forward estimates and back estimates can be made
- Fault occurrences as well as failure occurrences are handled
- Software is developed to allow efficient application



Shuttle Dynamic Failure Probability Evolution
Accounting for Failure-Free Tests plus Fault and Failure Occurrences and Corrections
Smoothed Mean Value
Failure Probability
14/100
12/100
10/100
8/100
6/100
4/100
2/100
0
1
21
41
61
81
101
121
Flight Seqence Number

Summary

- Key problems for a new system are the number of tests to conduct and the tracking of reliability
- The analysis needs to incorporate engineering information, reviews and oversights, and statistical data
- Bayesian analysis has these capabilities for dynamically updating estimates and quantifying uncertainties
- The application to number of tests needed shows the importance of incorporating the initial assurance level
- The application in tracking Shuttle shows the importance of dynamically tracking actual reliability growth

Annotated References (1)

1. Gaver, D.P., and Jacobs, P.A., Testing or Fault Finding for Reliability Growth: A Missile Destructive-Test Example, Naval Research Logistics, Vol. 44 (1997), pp. 623-637 (Applies also to tests or flights with fault identification and correction.)
2. Sen, Ananda and Gouri, Bhattacharyya, A Reliability Growth Model Under Inherent and Assignable-Cause Failures, Balakrishnan, N. Recent Advances in Life-Testing and Reliability, CRC Press, 1995, pp. 295-311 (Statistical approach for separating faults and failures according to cause)
3. G. Glenn Shirley, A Defect Model of Reliability, 1995 International Reliability Symposium, Available at http://web.cecs.pdx.edu/~cgshirl/Glenns%20Publications/31%201995%20A%20Defect%20Model%20of%20Reliability%20IRPS95%20Tutorial%20Slides.pdf (Defect model applied to yield defects but methodology is detailed and generally applicable)

Annotated References (2)

4. Fenton, Norman, Neil, Martin, and Marquez, David, Using Bayesian Nets to Predict Software Defects and Reliability, Available at www.agenarisk.com/resource/white.../fentonMMR_Full_vi_0.pdf (Applicable to hardware defects also. Useful tool for aggregating information)
5. Walls, L.A. and Quigley, J.L., Building Prior Distributions to Support Bayesian Reliability Growth Modeling Using Expert Judgment, Reliability Engineering and System Safety, Vol. 74(2), (2001), pp.117-128 (Useful approach for assessing initial assurance level)
6. Li, Guo-Ying, Wu, Qi-Guang, and Zhao, Yong-Hur, Bayesian Analysis of Binomial Reliability Growth, Journal of Japan Statistics, Vol 32, No. 1 (2002), pp.1-14 (Applicable to both fault and failure counts)
7. ESAS, Final Report, NASA-TIM-2005-214062, November 2005, Chapter 8. Risk and Reliability, Appendix 8C. Reliability Growth (Reliability growth models used in NASA's ESAS study)
8. Kececioglu, Dimitri, Reliability Engineering Handbook, Volume 2, Simon and Schuster, 1991, Chapter 16. Reliability Growth (Basic reliability growth models)

Perspective on Planetary Entry, Descent, and Landing Research

Contributions and Lessons Learned

Sean A. Commo

NASA Langley Research Center

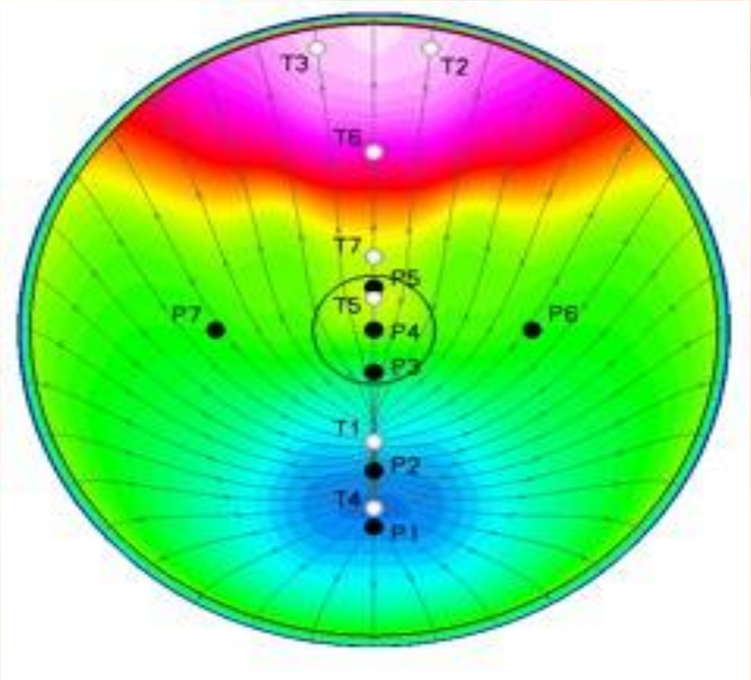
Engineering Directorate

05 May 2011

Introduction

- Entry, descent, and landing (EDL) is the phase of flight defined from atmospheric interface to touchdown on a planetary surface
- Future planetary missions strive to deliver larger payloads at higher altitudes with increased landing accuracy; currently driven by available EDL technologies
  - Nearing the limit with current technologies for Mars Science Laboratory (MSL); currently scheduled to launch Fall 2011
  - Stringent requirement for any manned missions; for Mars: two order of magnitude increase in landed payload mass, four order of magnitude increase in landing accuracy
- Current EDL systems are based on Viking-era technologies (1970's NASA Mars program) with minor modifications
  - Aeroshell geometry, thermal protection system (TPS) material, parachute design
- Development of newer, more robust EDL technologies rely on improving Earth-based modeling capabilities
  - Large uncertainties in computational simulations
  - Inadequate ground-based testing facilities
  - Sparse amount of flight data available

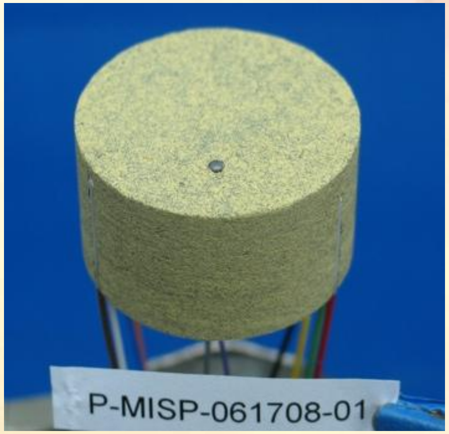
MEDLI Overview



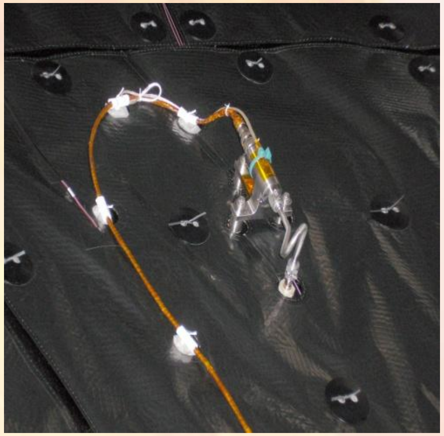
- The MSL Entry, Descent, and Landing Instrumentation (MEDLI) is a suite of sensors installed on the forebody heatshield of the MSL entry vehicle
  - Sensor locations determined by science team
  - Some similar components to previous entry instrumentation packages
- MEDLI operational from ten minutes prior to atmospheric interface to heatshield separation
- MEDLI proposed to address some of the challenges associated with development of newer, more robust EDL technologies
- MEDLI High-Level Objectives:
  - Provide more than an order of magnitude more data than all previous Mars entry missions combined
  - Answer fundamental questions relating to leeside turbulent heating levels, forebody flow transition, and TPS material response in a carbon dioxide atmosphere
  - Permit a more accurate post-flight trajectory reconstruction
  - Allow separation of aerodynamic and atmospheric uncertainties in the hypersonic and supersonic flow regimes.

MEDLI Subsystems

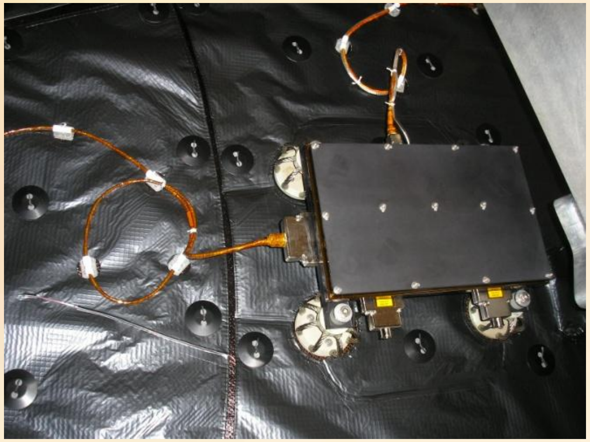
- MEDLI Integrated Sensor Plugs (MISP)
  - A plug consists of a 1.4” diameter heatshield TPS core with embedded thermocouples and recession sensors
  - Each plug consists of one (1) recession sensor and four (4) thermocouple sensors
  - Supports aerothermodynamic and TPS science objectives
- Mars Entry Atmospheric Data System (MEADS)
  - Series of through-holes, or ports, in the TPS that connect via tubing to pressure transducers
  - Based on Shuttle Entry Air Data System (SEADS)
  - Supports aerodynamic and atmospheric science objectives
- Sensor Support Electronics (SSE)
  - Electronics box that conditions sensor signals and provides power to MISP and MEADS



MISP



MEADS Pressure Transducer



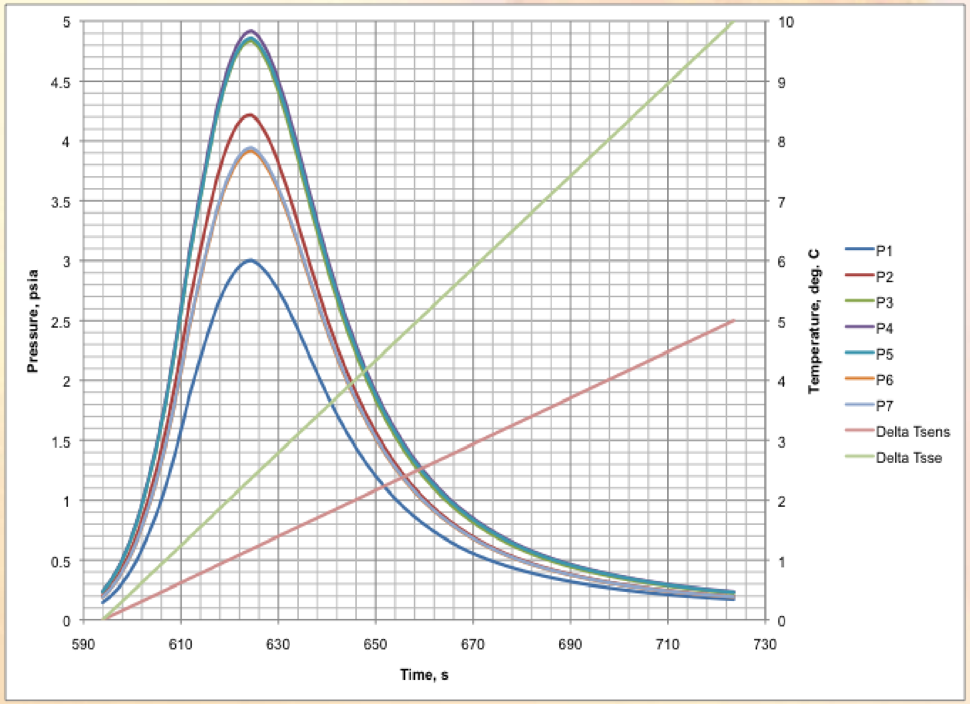
SSE

Pressure Measurement System

- Components of the pressure measurement system: seven (7) MEADS pressure transducers, SSE

- Pressure Measurement System Science Objectives:
  - Estimate flight parameters from measured pressures
  - Improve atmospheric models (density) for Earth-based computational simulations

- Defendable uncertainty in estimated flight parameters rely on adequate measurement system characterization over extreme environmental conditions



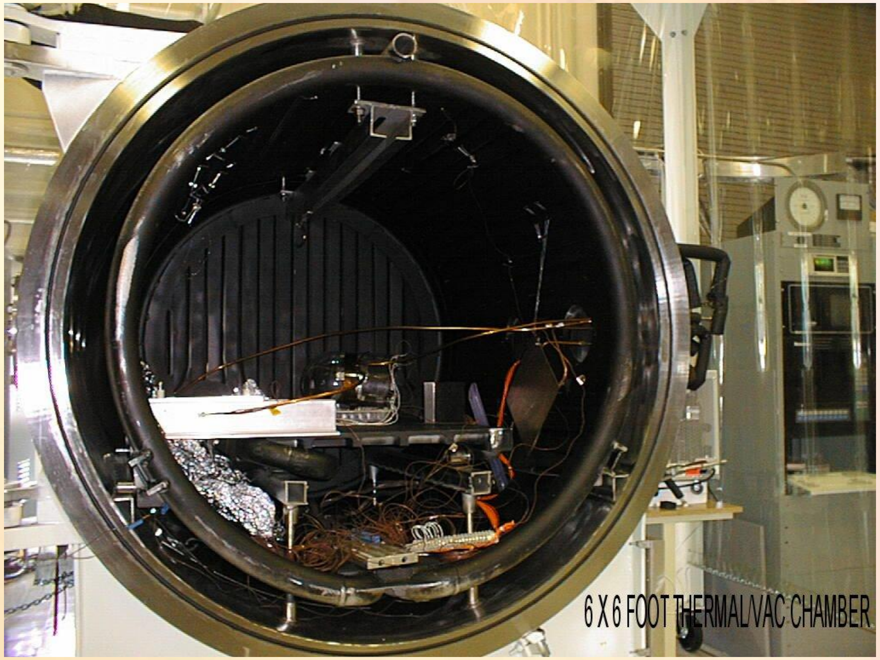
- Pressure Measurement System Characterization Challenges:
  - Pressure varies across port locations on the heatshield; temperatures vary between the SSE and pressure transducer locations; pressure and temperature vary with time during reentry
  - Possible Operational Ranges: 0.00 to 5.00 psia (Pressure), -120 to -60 deg. C (Transducer Temperature), -20 to 55 deg. C (SSE Temperature)
  - Large temperature ranges represent the uncertainty in the *start* temperatures (i.e. transducer temperature can start anywhere in the range of -120 to -60 deg. C with an expected change of 10 deg. C over the entry)

Pressure Measurement System Characterization - I

- Objective: Adequate measurement system characterization (calibration) over extreme environments; deliverables include
  - Mathematical model to estimate flight pressures
  - Uncertainty estimates throughout the flight trajectory
    - Total measurement uncertainty goal of 1 percent of reading through the range of 0.12 to 5.00 psia
- Utilized response surface techniques to provide a robust, defendable system characterization
  - RSM-based calibrations have been performed at NASA LaRC since 1999 (force balance applications)
  - Nontraditional use of RSM: not interested in system optimization; deliverables are measurement system knowledge (mathematical model and uncertainties)
  - Certain experimental design properties are important to providing a robust mathematical model that can be applied confidently to flight data
- Experimental Design Development
  - Mathematical model based on second-order Taylor series expansion of three factors
  - Replication included to estimate the pure experimental error in the measurement system (one metric for comparison of transducers)
  - Prediction variance properties of the design translate to total measurement uncertainty of the system

Pressure Measurement System Characterization - II

- NASA LaRC 6' x 6' Thermal Vacuum Facility
  - Provides the necessary testing conditions to characterize the pressure measurement system based on possible environmental conditions
  - Limitations of the Facility
    - The pressure measure system can stabilize with temperature within 2 hours and with pressure within 1 minute
    - Restrict the randomization of temperature to improve the experimental efficiency (split-plot)
      - Temperature combination is set and held constant while the pressure levels are varied



- Impact of Restricted Randomization
  - Since temperature is held constant while a pressure sequence is executed, there is a degree of correlation among the points; however different temperature combinations are independent
  - Require some advanced technique to perform the analysis which accounts for the restricted randomization
    - Restricted maximum likelihood (REML)
- Statistical calibration problem: develop forward regression model and invert to solve for estimated parameter

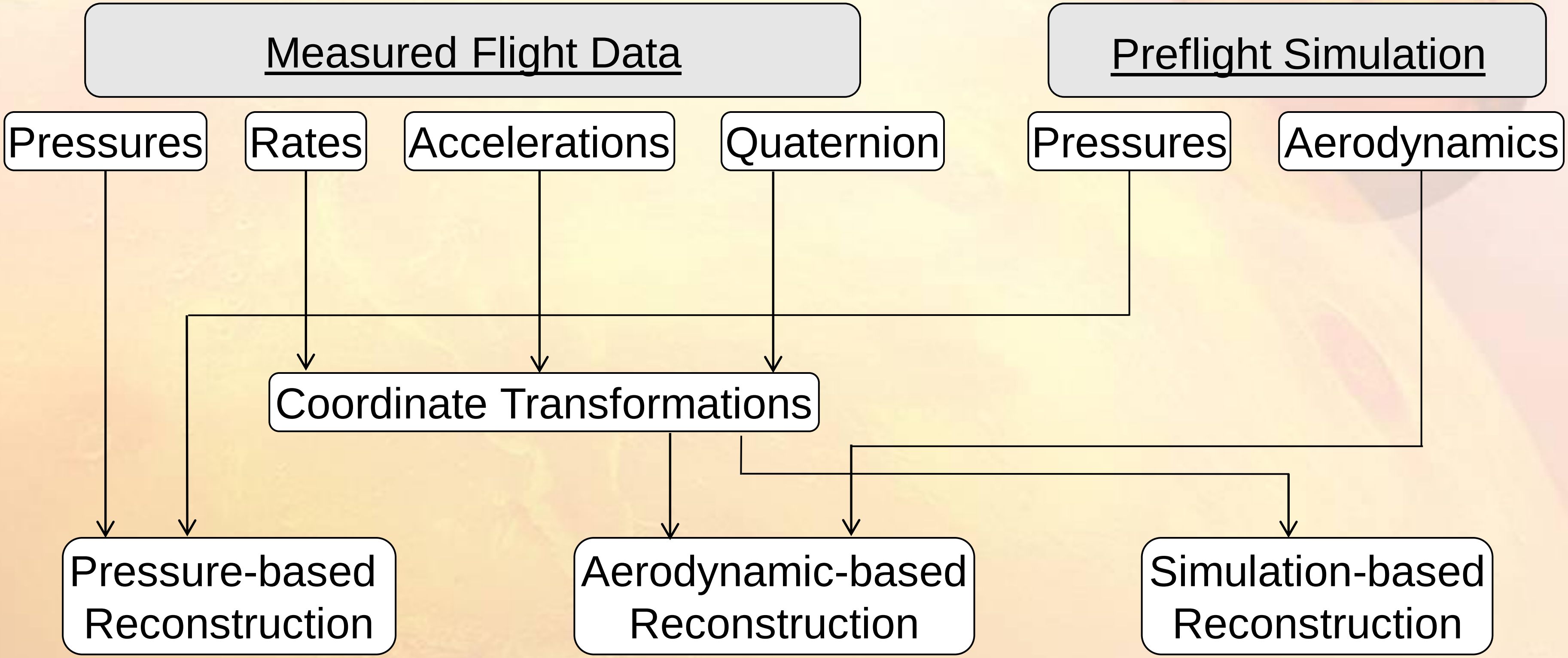
$$V = f\left(P, T_{\text{sens}}, T_{\text{SSE}}\right) \qquad \Rightarrow \qquad P = f\left(V, T_{\text{sens}}, T_{\text{SSE}}\right)$$

Pressure Measurement System Characterization Summary

Methodology and Tools Overview:

1. DOE/RSM:
   - Development of the experimental design to support objectives
   - Accommodate practical restrictions (restricted randomization)
   - Simulated entry trajectories: best attempt to simulate expected flight conditions on the ground
2. Transducer Repeatability:
   - Pure error estimation
3. Forward Regression Modeling:
   - REML: variance component estimation and model reduction
4. Model Inversion for Flight Data Reduction:
   - Estimate pressure from signal response and temperatures
   - Two (2) methods available: direct or iterative
5. Inverse Prediction Uncertainty:
   - Delta method used to calculate the variance in the estimated pressure
6. In-flight Zero Algorithm:
   - Exploit known, physical information prior to entry (hard vacuum in space)

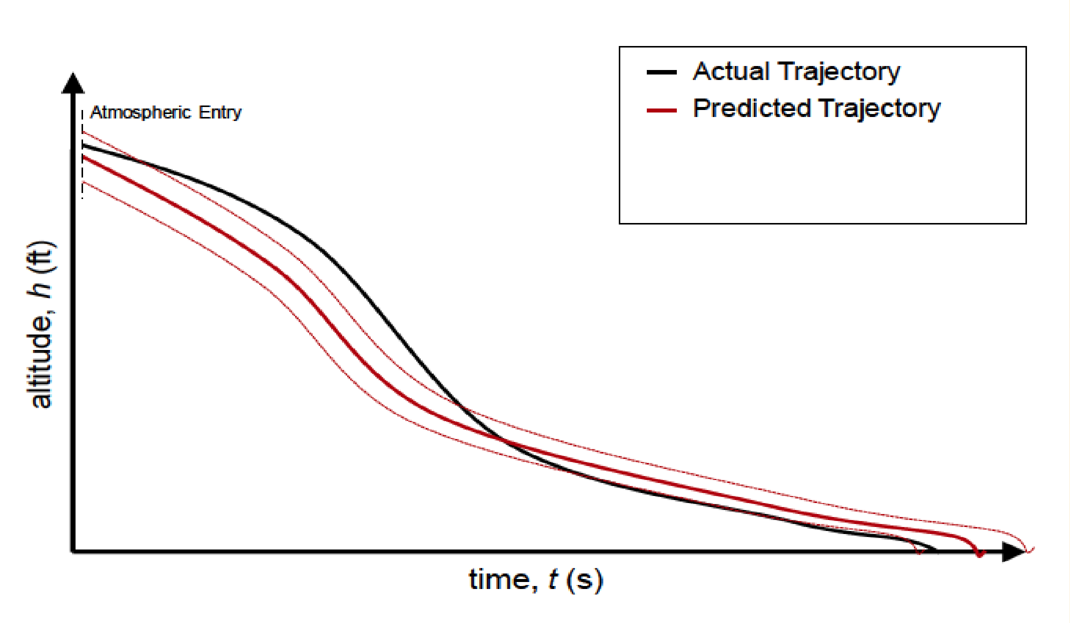
Post-Flight Trajectory Reconstruction



- Three (3) quasi-independent methods to reconstruct the entry trajectory
  - Ideally all the reconstructed trajectories match
  - Historically discrepancies have existed in the reconstructed trajectories which have not been systematically resolved
- System-level approach to quantifying uncertainties has not been emphasized for previous reconstruction efforts
- Emphasize more strategic approach to help meet objectives: Monte Carlo vs. Designed Experiment

Reconstruction Reconciliation

- Research effort with Georgia Tech (Jason Corman and Brian German) under the funding auspices of the NASA Graduate Student Research Program (GSRP)
- Focused on the development of a general approach to determine the causes in differences between various trajectory reconstruction methods
  - Reconstructed trajectories do not need to match exactly
  - Uncertainty intervals of the trajectories to overlap
- Developed a simplified, 2 degree-of-freedom (DOF) simulation tool to study trends and sensitivities
  - Identified and tested techniques to help mitigate discrepancies in basic reconstruction methods
  - Apply the approach to the actual 6 DOF simulation tool used during reconstruction



Pressure-based Trajectory Reconstruction

- Combines actual flight pressure data and preflight simulation data to estimate vehicle orientation, freestream dynamic pressure, and Mach number
- Preflight simulation data based on computational fluid dynamics (CFD) with limited anchoring to experimental data
  - Experimental facilities available are not relevant to expected flight environment
  - Higher confidence in computational results in certain regions of the trajectory
- Uncertainty requirements for estimated flight parameters (angle of attack, angle of sideslip, Mach number, dynamic pressure) specified at project's inception
- Uncertainty requirements were determined assuming perfect (no uncertainty) preflight simulation data
  - Investment of resources focused on minimizing the uncertainty in the pressure measurement uncertainty
- Total uncertainty is the root sum squared (RSS) of the pressure measurement system uncertainty and the preflight simulation uncertainty
  - Pressure Measurement System Uncertainty ~ ± 0.25 percent (actual)
  - Preflight Simulation Uncertainty ~ ± 5 percent

$$\text{Total Uncertainty} = \sqrt{(0.25)^2 + (5)^2} \approx 5$$

Summary

- Development of more robust EDL technologies rely on capabilities of Earth-based modeling capabilities
- Contributions to MEDLI:
  - Pressure Measurement System Characterization
    - Mathematical modeling of the pressure measurement system
    - Defendable uncertainty quantification of the system
  - System-level Approach to Preflight Trajectory Reconstruction
    - Subsystem uncertainty quantification
    - General approach to reconstruction reconciliation
- Lessons Learned from MEDLI:
  - Traceable objectives to support future missions
    - Development of technologies
    - Decision-making process
  - Investment of resources to support objectives
    - Division between computational and physical experiments
  - Focus on knowledge and learning rather than what needs to be done or built
    - De-emphasizes data quantity

References

- Some Textbooks
  - Draper, N.F. and Smith, H. (1998), *Applied Regression Analysis*, John Wiley & Sons
  - Myers, R.H. and Montgomery, D.C. (2002), *Response Surface Methodology*, (2nd Ed.) John Wiley & Sons.
  - NIST Engineering Statistics Handbook, *Measurement Process Characterization*, http://www.itl.nist.gov/div898/handbook/mpc/mpc.htm
- Some Articles, annotated
  - Braun, R.D. and Manning (2006), "Mars Exploration Entry, Descent, and Landing Challenges," *Journal of Spacecraft and Rockets*, 44, pp. 310-323, March 2007.
  - Gazarik et al. (2008), "Overview of the MEDLI Project," *2008 IEEE Aerospace Conference*, Big Sky, MT.
  - Kowalski, S.M., Parker, P.A., and Vining, G.G. (2007), "Tutorial on Split-Plot Experiments," *Quality Engineering*, 19, pp. 1-15.
  - Parker et al. (2010), "The Prediction Properties of Classical and Inverse Regression for the Simple Linear Calibration Problem," *Journal of Quality Technology*, 42, pp. 1-16.
  - Parker, P.A. and DeLoach, R. (2001), "Response Surface Methods for Force Balance Calibration Modeling," *IEEE 19th International Congress on Instrumentation in Aerospace Simulation Facilities*, Cleveland, Ohio.

Thermal and Pressure Characterization of a Wind Tunnel Force Balance using the Single Vector System (SVS)

(AIAA-2011-950)

Authors: Chris Lynn, Sean Commo, Tom Johnson, Pete Parker
NASA Langley Research Center

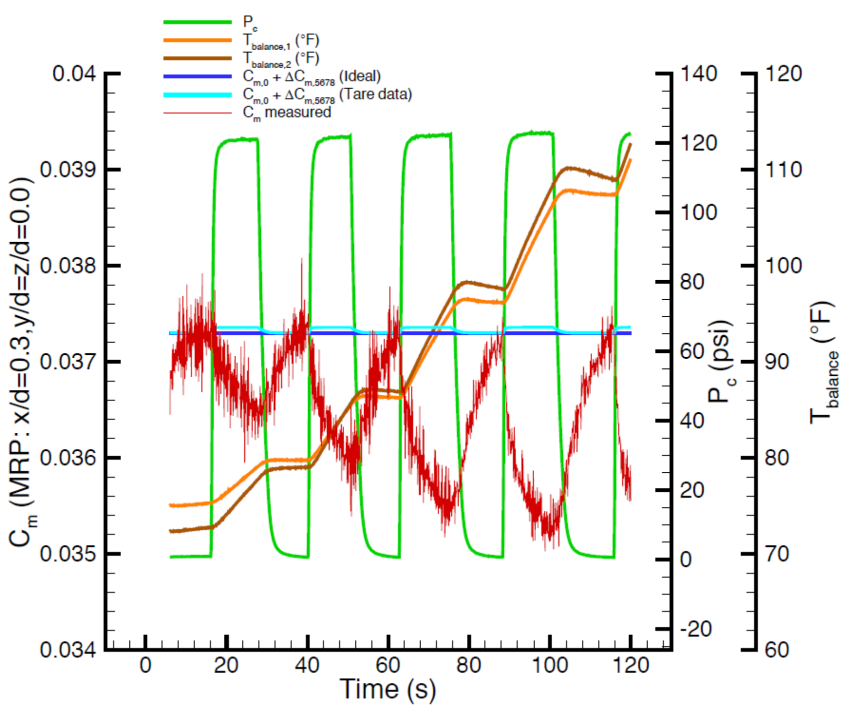
Chris Lynn
NASA Langley Research Center
Aeronautics Systems Engineering Branch (ASEB)
National Force Measurement Technology Capability (NFMTC)
May 5, 2011

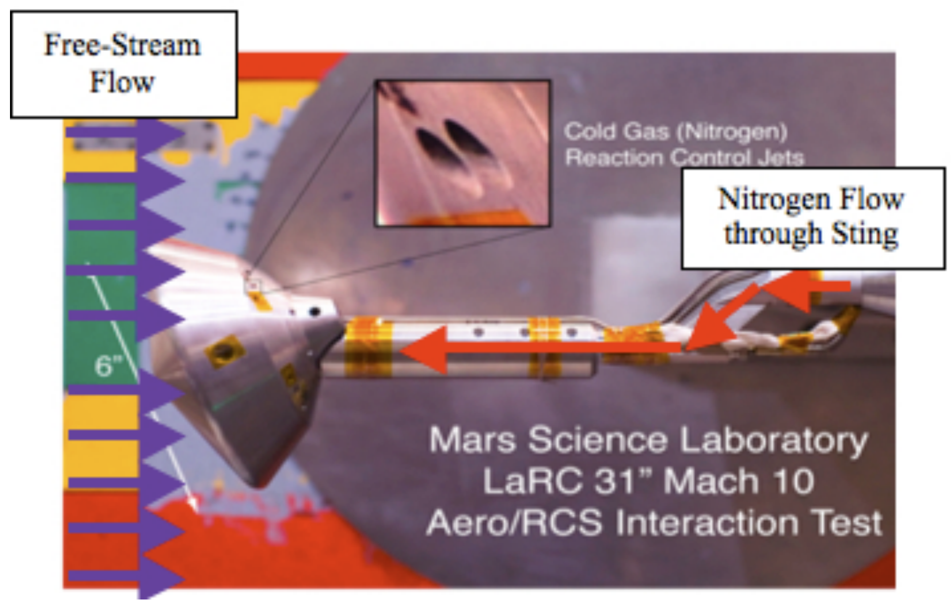
Overview

- Introduction
- Motivation – MSL Test Objectives
- Calibration Techniques
- SS-12 Force Balance Design
- Experimental Design
  - Pre-Planning
  - Experimental Design (Load Schedule) Development
    - Crossed Design
    - IV Optimal Design
  - Execution Strategy and Implications
  - Experimental Design Properties
- Calibration Setup & Execution
- Data/Error Analysis & Model Comparison
- Conclusions

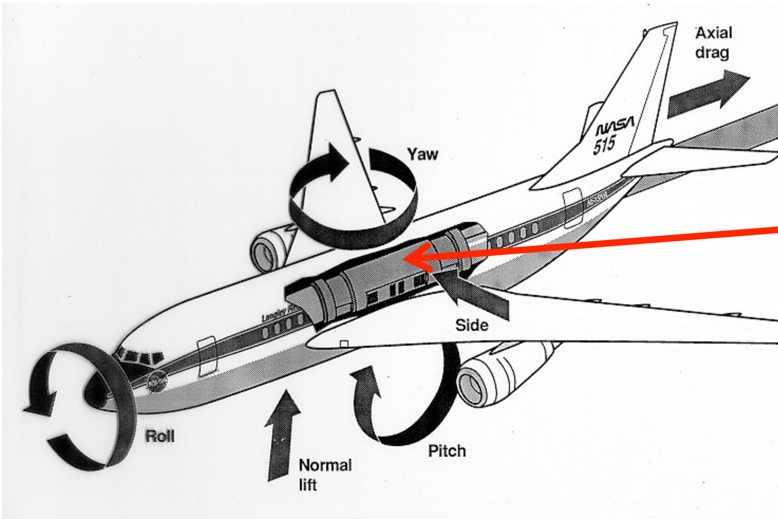
Motivation

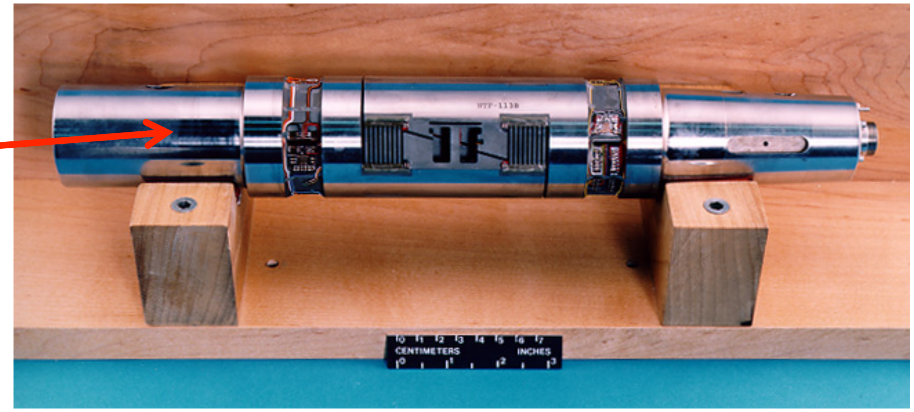
- Previous testing at the NASA LaRC 31-Inch Mach 10 facility with the Mars Science Laboratory (MSL) aeroshell revealed several thermal related issues during tests.
- Primary issue involved temperature drift of the force balance over the duration of each blow-down run.
- Worst-case temperature drift observed during RCS pressure cycle runs (cycling reaction control jets on-off) was ~ 40°F over 120 second run time.
- MSL research team proposed the following problem statement:
  - Pursue having balance team at NASA LaRC design method for characterizing the outputs of the strain-gages subject to various forces, moments, pressures, and temperatures





What is a Force Balance?

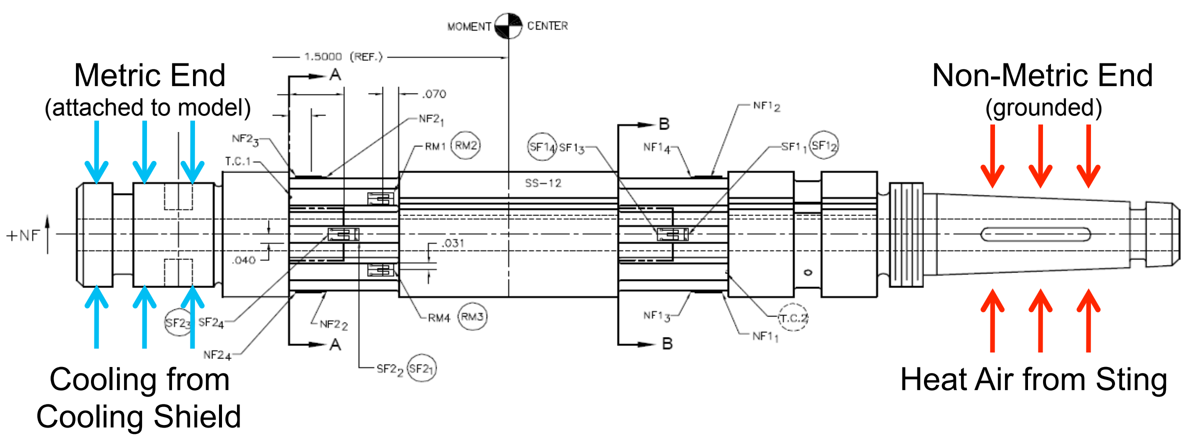




- Force Balances are transducers used to measure the 6DOF aerodynamic loads encountered by a wind tunnel model during a wind tunnel test
- Balances are complex structural spring elements composed of flexural elements – only structural component between model (metric end) and sting (non-metric end)
- Flexures are instrumented with foil resistive strain gages that output an electrical signal which is proportional to the strain level induced onto the flexural elements
- Balance structure and instrumentation designed to be sensitive to only single component applied loads/moments, but imperfections (in machining/instrumentation) require us to characterize interactions

SS-12 Force Balance

- SS-12 is a single-piece, 5-component, water-cooled, flow-thru force balance (no Axial Force)
- Balance has a concentric hole down its center, allowing flow of gas thru balance and out to attached model on metric end
- Normal force and side force components re-gauged from direct read to a force type balance configuration (N1, N2, S1, S2, RM)
  - N1/N2, S1/S2 bridges at a single axial station along balance
  - Single station bridging of strain gages aids in reducing sensitivity of measurement bridges to thermal effects
- Balance design features an active cooling shield that covers balance during use, and actively cycles water around balance (typical for balances used in supersonic/hypersonic testing regimes)



Component	Design Load
NF	100 lbs
AF	n/a
PM	150 in-lbs
RM	32 in-lbs
YM	40 in-lbs
SF	30 lbs

Pre-Experimental Planning

Collaboration amongst all colleagues to clearly establish objectives and factor level settings.

- A clear statement of the goals/objectives of an experiment is critical (Answer the right question(s).)
  - Objective:
    1. Characterize the outputs of the strain-gages subject to various forces, moments, pressures, and temperatures (develop continuous functions for each measurement component)
- Selection of the factors and measured responses
  - Aerodynamicists, Force Measurement Engineers and Statistical design experts collaborated to determine optimal solution to meet objectives
  - Design, Held-Constant, Uncontrolled Factors
  - Design Factor ranges:
    - Forces and Moments → full-scale range balance design loads (which match the expected test loads)
    - Pressure and Temperature → range over expected operating conditions during wind-tunnel test
  - Design Factor Levels: support experimental objectives

Design Factors:

Factor Label	Design Factor (units)	Range
A	Normal Force *(lbs)*	-100 to +100
B	Pitching Moment *(in-lbs)*	-150 to +150
C	Rolling Moment *(in-lbs)*	-32 to +32
D	Yawing Moment *(in-lbs)*	-40 to +40
E	Side Force *(lbs)*	-30 to +30
F	Average Balance Temperature*(°F)*	70 to 120
G	Balance Cavity Pressure *(psia)*	14.7 to 400

Measured Responses:

Response	Response Type (units)
1	Normal Force Bridge Output *($\mu V/V$)*
2	Pitching Moment Bridge Output *($\mu V/V$)*
3	Rolling Moment Bridge Output *($\mu V/V$)*
4	Yawing Moment Bridge Output *($\mu V/V$)*
5	Side Force Bridge Output *($\mu V/V$)*

Experimental Design

- Fundamentals of Statistical Design of Experiments
  - Randomization: defends against systematic errors (i.e. hysteresis) in an experiment.
  - Replication: provides information on the pure experimental error in the response, which sets the lower bound for uncertainty.
  - Blocking: limits the effects of any nuisance (controlled or uncontrolled) factors in an experiment.
- Postulated Mathematical Model: based on Taylor series expansion
- Balances are highly dimensional instruments, requiring response surface methods to properly characterize performance over design-space
- Two Experimental Designs Generated & Executed
  - Crossed Design & IV Optimal Design (both designs are Split-Plot (SP) designs)
  - Forces/Moment load schedules based off Central Composite Design (CCD)

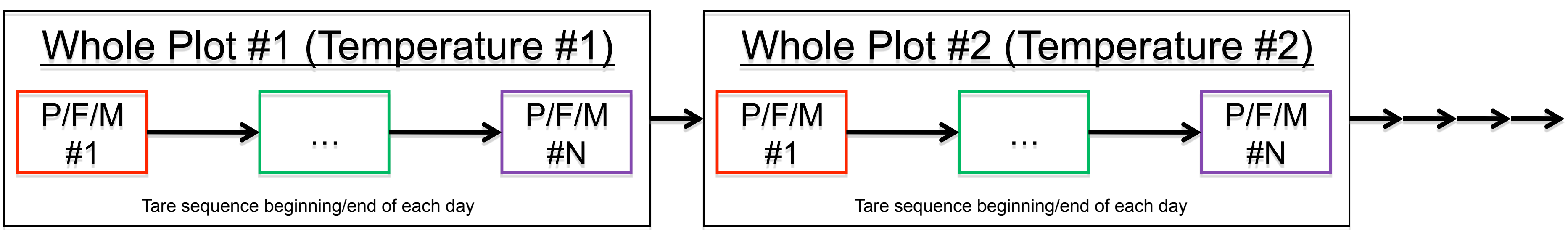
$$y = \beta_0 + \sum_{i=1}^{7} \beta_i x_i + \sum_{i=2}^{7} \beta_{ii} x_i^2 + \sum_{i=1}^{7} \sum_{j=i+1}^{7} \beta_{ij} x_i x_j$$

$$x_1 = \text{Temperature} \quad x_2 = \text{Pressure} \quad x_3 = \text{NF} \quad x_4 = \text{PM} \quad x_5 = \text{RM}$$

$$x_6 = \text{YM} \quad x_7 = \text{SF}$$

Calibration Execution *(Crossed Design)*

- Presence of hard-to-change factors in an experiment can make a completely randomized experimental design impractical to implement.
  - Temperature is often an expensive/time-consuming factor to change
- Split-Plot designs are a technique to deal with experiments with hard-to-change factors
  - Restrict randomization for hard-to-change factors
  - Concept developed from agricultural experiments
- Temperature is set and held constant while pressure, forces, and moments combinations are varied randomly for each point within each whole plot
  - Because of time required to complete, design is blocked by day. Once temperature is set, it does not change for the rest of that day.
- Calibration occurred over the course of 8 days
  - 2 days for standard calibration
  - 6 days for pressure/temperature calibration



Development of the Experimental Design 1
(Crossed Design)

5-Component SVS Design:

Factorial Design Points

NF	PM	RM	YM	SF
-48	-72	± 15	-19	-13
-48	-72	± 15	19	13
-48	72	± 15	-19	13
-48	72	± 15	19	-13
48	-72	± 15	19	-13
48	-72	± 15	-19	13
48	72	± 15	-19	-13
48	72	± 15	19	13
0	0	0	0	0

Axial Design Points

NF	PM	RM	YM	SF
± 100	0	0	0	0
± 100	± 150	0	0	0
± 100	0	0	0	± 30
± 100	0	± 32	0	0
0	0	0	± 40	± 30
0	0	0	0	± 30
0	0	0	0	0

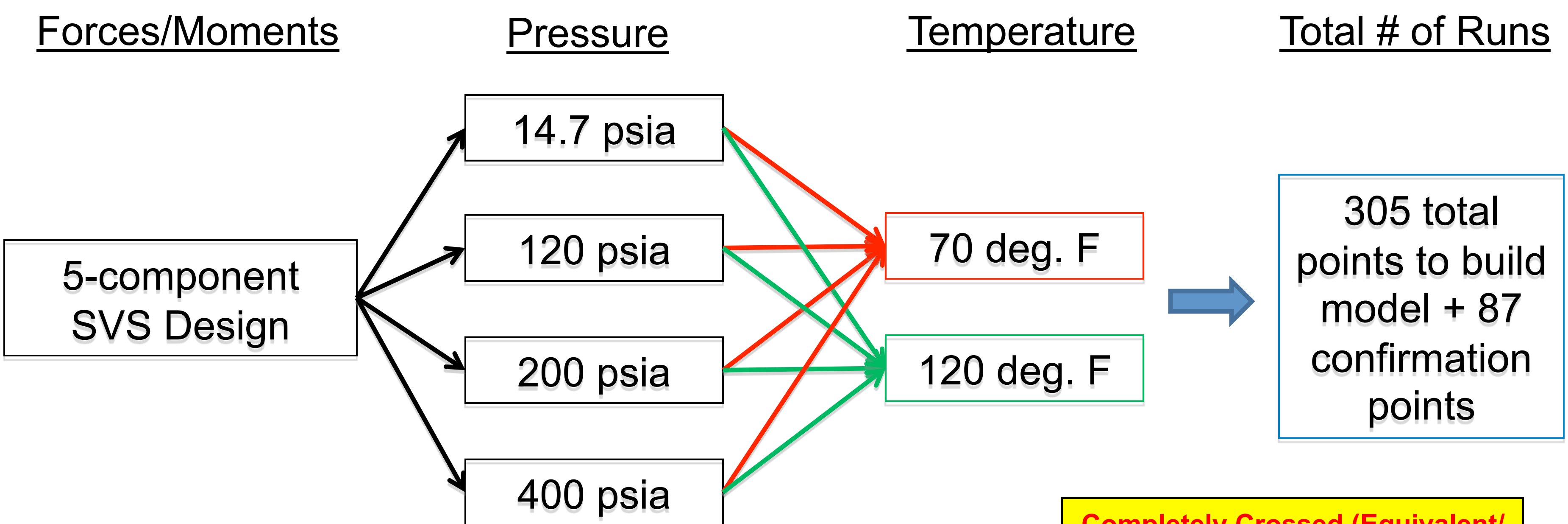
Factorial Points:	16
Axial Points:	20
Center Points:	6
Total:	***42***

Development of the Experimental Design I
(Crossed Design)

Mathematical Model Assumptions:

- First-order effect of temperature on responses (req. 2 unique levels)
  - Assumption based on experience, historical data
- Second-order effect of pressure, forces, and moments on responses (req. 3 unique levels)

Experimental Design Development:



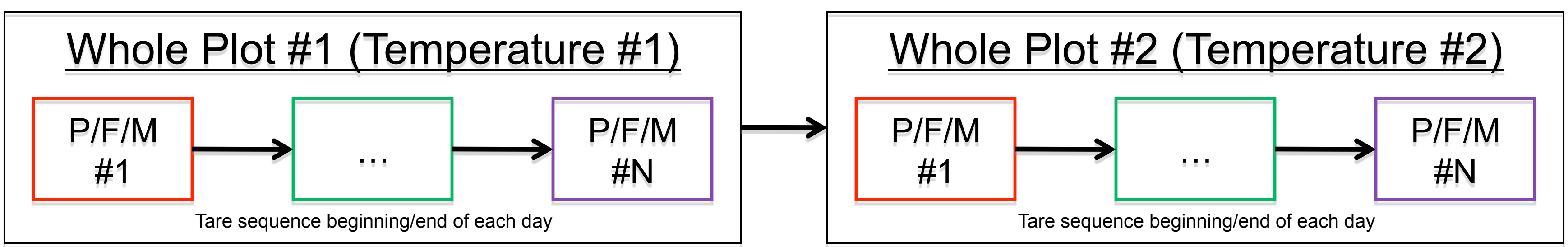
Completely Crossed (Equivalent/ Balanced Design)

Development of the Experimental Design II & Execution *(Optimal Design)*

Mathematical Model Assumptions:

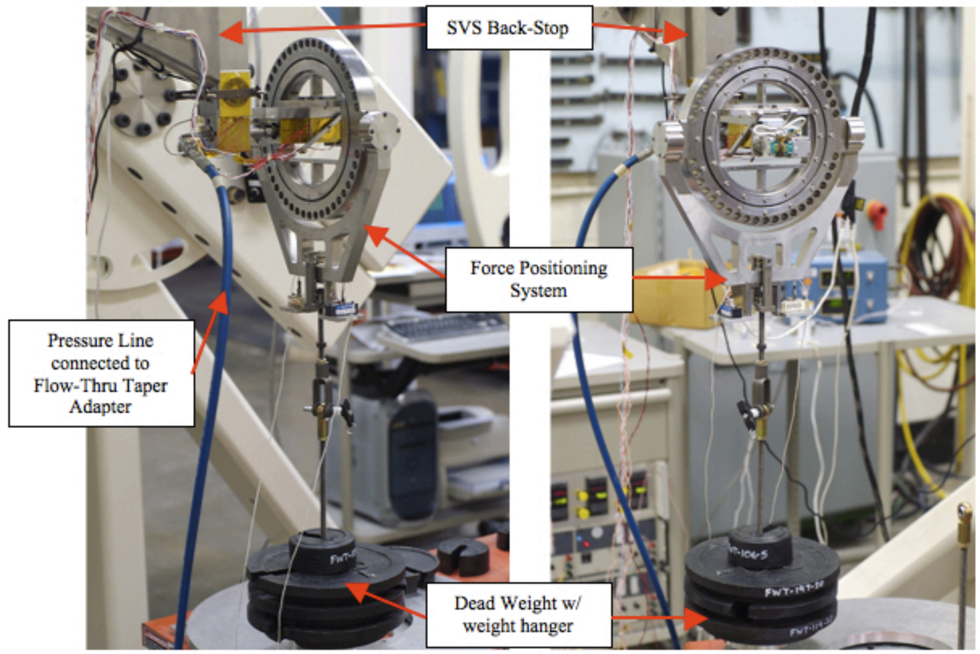
- Same as Crossed Design

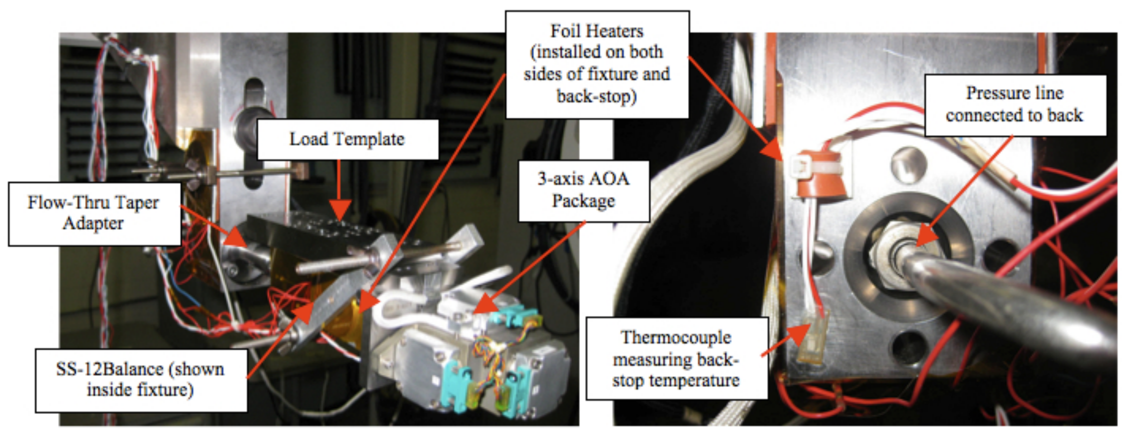
- Crossed experimental design points used as a candidate list within DE to generate an optimal design
  - Completely Crossed design contained all possible combinations possible using SVS, based on common CCD design
  - IV Optimal Design used as it provides a lower prediction variance across the design space (desirable for instances when prediction capability is critical)
  - 44 point design generated & executed (plus 42 point room temp design – needed to provide sufficient DOF in order to compute T main effect term)
- Design properties (leverage, VIF's, SE) inspected to ensure good selection of design points



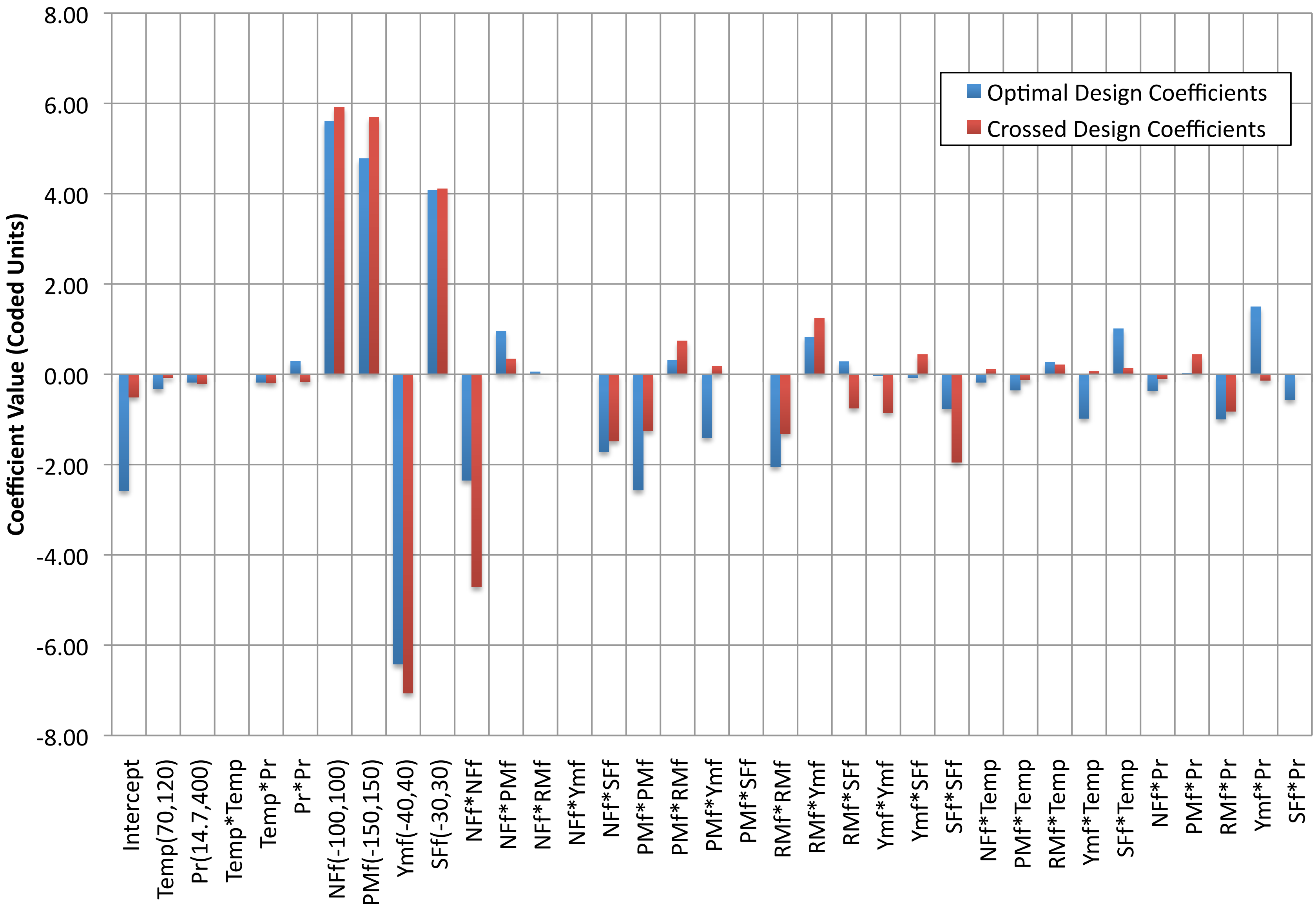
Calibration Setup

- Single Vector System (SVS) used during calibration to orient balance
- Heater system configured with foil heaters on balance calibration block and SVS back-stop to elevate steady balance temperature to desired settings.
  - Temperatures actively controlled to within 1-2 °F of desired set point
- Static pressure applied to internal balance cavity via pressure fitting in rear of stump adapter.
  - Cap plate on forward most end of balance calibration block sealed off system, allowing application of static pressure to balance
  - Nitrogen k-bottle used for air source

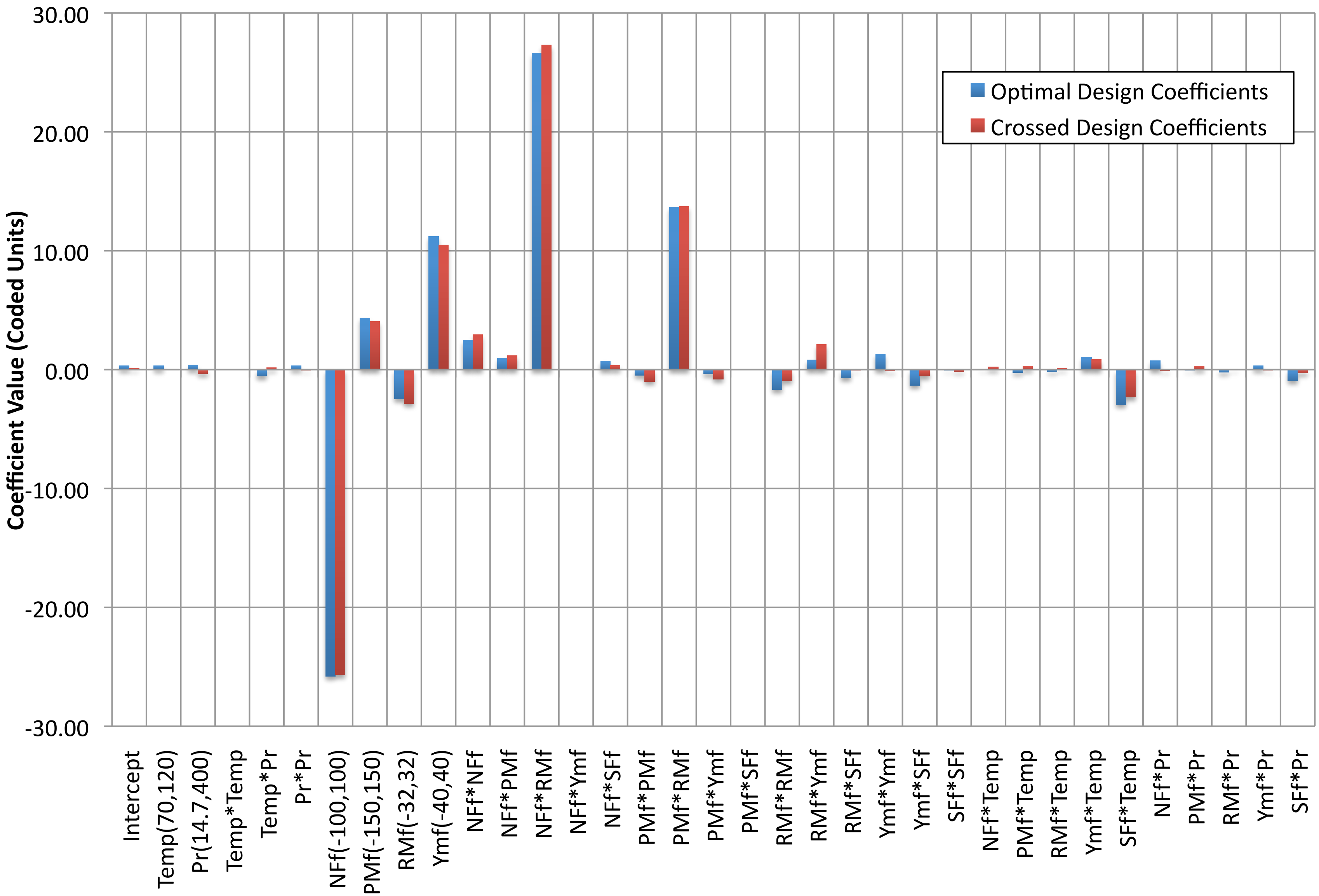




Results: Model Comparison
(RM Response)



Results: Model Comparison
(SF Response)



Results: Model Summary

- No model reduction employed
- Tare data collected during calibration used to reduce data, in order to get total applied loads
- Data from both Completely Crossed & Optimal designs analyzed using REML in JMP
- Each design clearly shows improved prediction accuracy when Pressure/Temperature model coefficients included
- Small differences in prediction accuracy estimates exist between crossed and optimal designs

	2sigma (%FSE)				
	NF	**PM**	**RM**	**YM**	**SF**
optimal + room temp data, w/ P and T	0.1104	0.0918	0.5715	0.1579	0.2054
optimal + room temp data, w/o P and T	0.1497	0.1186	0.5404	0.1857	0.2503

	2sigma (%FSE)				
	NF	**PM**	**RM**	**YM**	**SF**
crossed design, w/ P and T terms	0.0849	0.0699	0.4637	0.1185	0.1628
crossed design, w/o P and T terms	0.1218	0.1077	0.4776	0.1645	0.2095

Conclusions

- An engineering problem was presented to the team. A methodical approach was developed to solve this problem, which combined efforts from both engineering and statistical fields of expertise.
- Demonstrated a method to characterize the force balance in the wind tunnel environment, including temperature and pressure, thereby improving aerodynamic research data quality
  - Data from calibration and MSL test data reveals significant improvement (multiple sources contribute to increased data improvement)
- Calibration data reveals both designs result in very similar mathematical models, with very similar residual/accuracy estimates
- Appropriate metrics were determined to evaluate the robustness of the experimental design, developed for this specific calibration.
- With appropriate planning and coordination, the methods described from this investigation can be applied to any calibration to yield a powerful mathematical model that characterizes the performance of the system under consideration.
- Resulting mathematical models (algorithms) generated were transferred to the wind-tunnel test team, and the on-board compensation techniques were applied real time during the test.

Contacts and References

- **Contacts for more information:**
  - Keith (Chris) Lynn (keith.c.lynn@nasa.gov)
  - Sean Commo (sean.a.commo@nasa.gov)
  - Thomas Johnson (thomas.h.johnson@nasa.gov)
  - Peter Parker (peter.a.parker@nasa.gov)
- **Some Textbooks:**
  - Montgomery, D.C., *Design and Analysis of Experiments, 7th Edition*, 2009, John Wiley & Sons, Inc., New York.
  - Myers, R.H. et. al., *Response Surface Methodology, 3rd Edition*, 2009, John Wiley & Sons, Inc., New York.
  - Montgomery, D.C. et. al., *Introduction to Linear Regression Analysis, 4th Edition*, John Wiley & Sons, Inc., New York.
- **Some Articles, Annotated:**
  - Schoenenberger, M. et. al., "Aerodynamic Challenges for the Mars Science Laboratory Entry, Descent, and Landing", AIAA 2009-3914, June 2009.
  - AIAA, "Recommended Practice: Calibration and Use of Internal Strain-Gage Balances with Applications to Wind Tunnel Testing", AIAA R-091-2003.
  - Parker, P.A., "A Single-Vector Force Calibration Method Featuring the Modern Design of Experiments", AIAA 2001-0170, January 2001.

Advancements in Aeronautics Measurement System Characterizations

Ray D. Rhew

NASA Langley Research Center

NASA Statistical Engineering Symposium

May 5, 2011

NASA's Aeronautics Test Program (ATP) is a model program created to preserve the capabilities of the largest, most versatile, and *comprehensive* set of testing facilities in the nation.

With NASA facilities located at the Ames Research Center (ARC) in Mountain View, California, the Glenn Research Center (GRC) in Cleveland, Ohio, and the Langley Research Center (LaRC) in Hampton, Virginia, the ATP offers government, corporations, and institutions a wide range of experimental test services that reflect 60 years of unmatched aerospace test history. The ATP maintains a nationwide team of highly trained and certified staff, whose backgrounds and education encompass every aspect of aerospace testing and engineering.

Regardless of the test requirements, NASA's ATP can provide its clients with test results of unparalleled superiority.

FACILITY SPECIFICATIONS

	Speed	Reynolds number, per foot
11-Foot Transonic Unitary Plan Facility	0.20 to 1.45 Mach	0.30 to 9.6×10^6
National Transonic Facility	0.1 to 1.2 Mach	4 to 146×10^6
Transonic Dynamics Tunnel	0.1 to 1.2 Mach	0.30 to 3×10^6 and 0.2 to 10×10^6
8-Foot High-Temperature Tunnel	4, 5, and 7 Mach	0.30 to 5.1×10^6
9- by 15-Foot Low-Speed Wind Tunnel	0 to 0.20 Mach	0 to 1.4×10^6
14- by 22-Foot Subsonic Tunnel	0 to 0.3 Mach	0 to 2.1×10^6
20-Foot Vertical Spin Tunnel	0 to 85 ft/s	0 to 0.15×10^6
4-Foot Supersonic Unitary Plan Wind Tunnel	1.5 to 2.9 and 2.3 to 4.6 Mach	0.5 to 6×10^6 and 0.5 to 11×10^6
Icing Research Tunnel	50 to 395 mph	—
10- by 10-Foot Supersonic Wind Tunnel	0 to 0.4 and 2.0 to 3.5 Mach	0.1 to 3.4×10^6 (closed-loop) and 2.1 to 2.7×10^6 (open-loop)
Aerothermodynamics Laboratory	6 and 10 Mach	0.05 to 0.7×10^6, 0.2 to 2.2×10^6, and 0.5 to 8.0×10^6
8- by 6-Foot Supersonic Wind Tunnel	0.25 to 2.0 and 0.0 to 0.1 Mach	1.7 to 4.8×10^6
9- by 7-Foot Supersonic Wind Tunnel	1.55 to 2.55 Mach	0.50 to 5.7×10^6

The goals of the ATP include:

- **increasing the probability of having the right capabilities in place at the right time**
- **operating facilities in the most effective and efficient manner possible**
- **to foster those capabilities through a corporate management philosophy**
- **ensuring intelligent investment and divestment while sustaining core capabilities**

- **Providing quality data (information) to answer key technical research and development questions**

Test Technology – Instrumentation

Critical Element of Data Quality

Force and angle-of-attack measurement technology are *key aeronautical capabilities* addressed by the National Aeronautics R&D Policy

- **"We will dedicate ourselves to the mastery and intellectual stewardship of the core competencies of Aeronautics", and "key aeronautical capabilities"**
- **Capability in these technologies are not ones that NASA can readily purchase - the instruments are complex and require an experience based competency**

ATP Established Test Technology Capability Projects

- **Establish a national test technology capability to support aeronautics test requirements for NASA, AEDC, and the nation**

- **Dedicate a few NASA engineers and technicians**
  - **Develop subject matter experts (SMEs) through *"hands-on" experience***
- **Centers/Programs cooperate to provide & fund FTE**
- **ATP/SCAP invest in maintenance and recapitalization projects**

Infusing Statistical Thinking

- **Developed strategic technical goals – One goal is to:**

- **Improve calibration/characterization and develop recommended practices – multi-component force and angle measurement systems lack traceable calibration system standards**

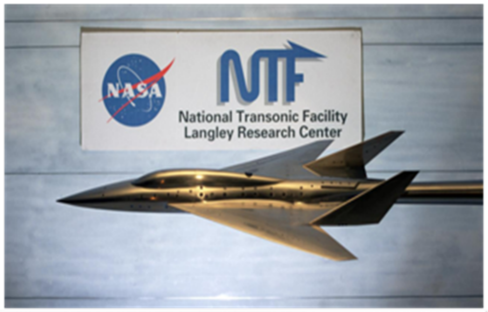
  - **Train SMEs on methodologies, tools and techniques**
  - **Bring statistical engineers into work hands-on projects to facilitate the training and improve knowledge transfer**

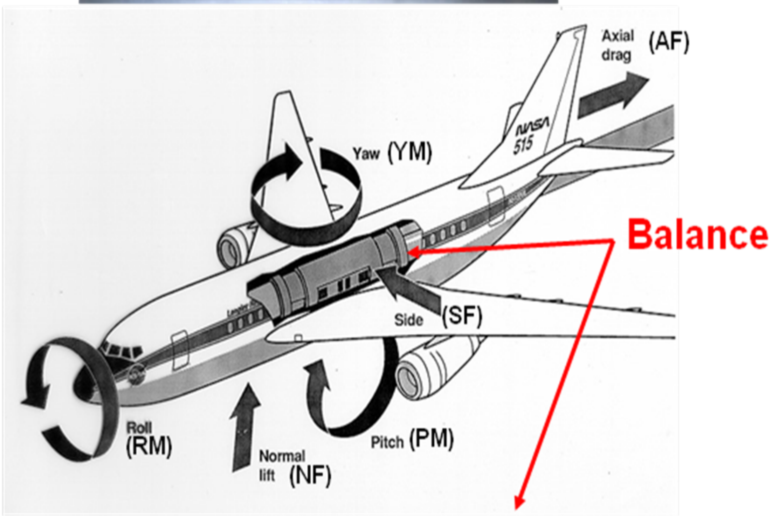
  - **Key: Continue questioning "WHY" – from the calibration systems to the experimental calibration designs to the calibration data analysis and model building**

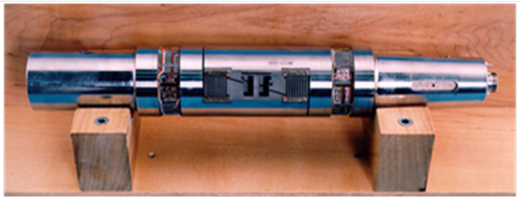
Calibration Goal: *Produce an accurate mathematical model (matrix) to estimate the aerodynamic parameters from measured responses.*

Force Measurement
Wind Tunnel Force Balances

- **Specialized force measurement instruments utilized in >90% of wind tunnel experiments**
- **Measure six components of load**
- **Highly stressed**
- **Failure is potentially catastrophic**







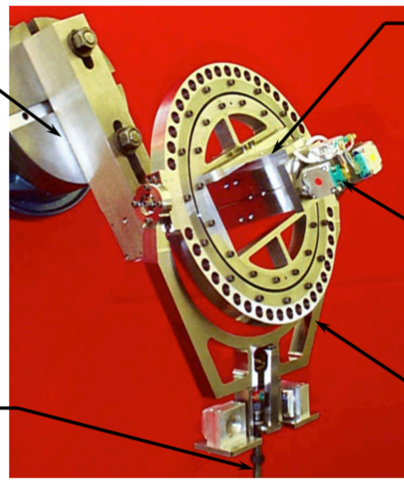
Wind Tunnel Force Balances

- **Calibration/Characterization Improvements**
  - **Implementation of statistical analysis techniques tailored for balances**
  - **Model building based on:**
    - **Statistical significance**
    - **Variance inflation**
  - **Results:**
    - **Improved models (more robust)**
    - **New information on uncertainty intervals**
    - **Insight into instrument behavior and calibration experiments**

BalFit Software Tool

Wind Tunnel Force Balances

- Calibration/Characterization Improvements
  - Implementation of the Single Vector System (SVS) as a standard technique
    - Integrates a Unique Force Application with DOE – Advantage of SME and Statistical Engineering collaboration (coupled experiment design and analysis into the calibration system development)
    - Randomization: to defend against systematic errors due to time, temperature or order of loading; Blocking: to prevent variation between blocks by breaking the experiment into manageable pieces; Replication: to cancel out random error and estimate pure error
  - Results:
    - <u>Increased Accuracy</u>, and <u>10x</u> <u>Reduction in Time and Cost</u>
    - Fewer sources of systematic error
    - Insight into instrument behavior and calibration experiments

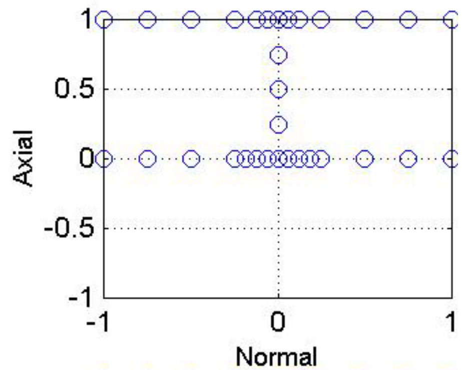


Wind Tunnel Force Balances

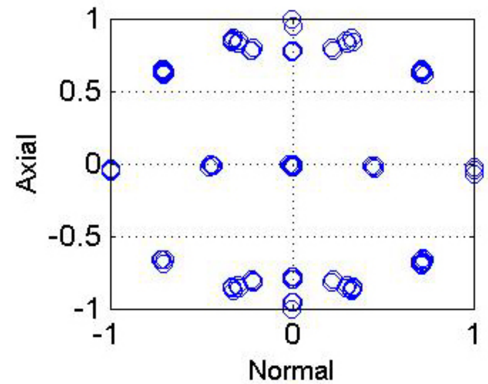
- **Calibration/Characterization Improvements**
  - **Implementation of design of experiments (DOE) techniques to more and complex calibrations**
  - **Adding additional factors such as temperature and pressure**

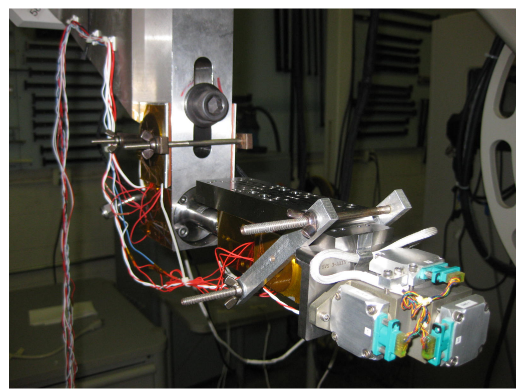
  - **Results:**
    - **Improved models (more robust)**
    - **More efficient experimental plans (number of runs)**
    - **Simpler implementation (more reasonable)**

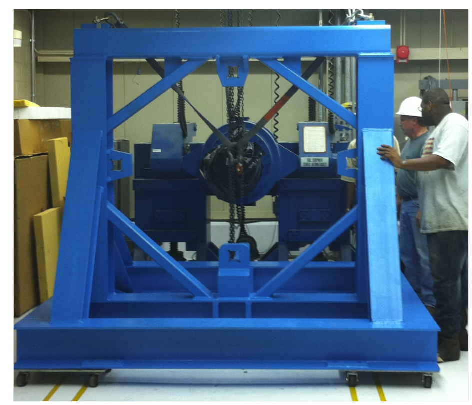
OFAT design



DOE design

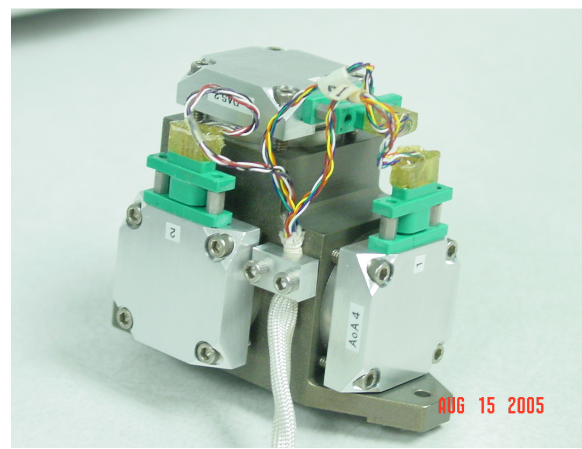






Angle Measurement Systems

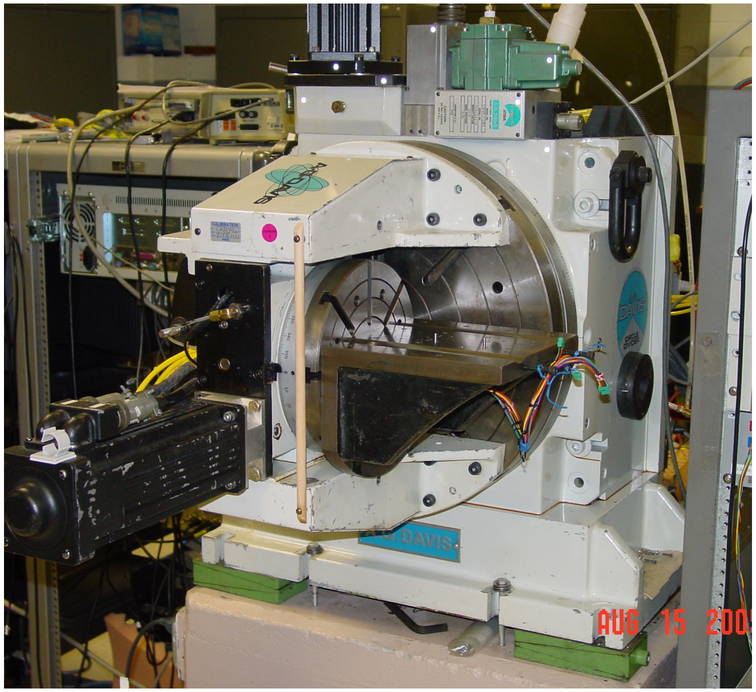
- **Angle Measurement System (AMS)**
  - **Utilized to determine the model orientation (Pitch and Roll)**
  - Three single-axis servo accelerometers (quartz flexures, Q-flex) mounted in a (near) orthogonal frame
  - Signal is proportional in magnitude and direction relative to earth's gravity vector
    - linearly proportional to the gravitational components
    - angle is proportional to the sine of the gravitational component



Tri-axial accelerometer (AMS)

Angle Measurement Systems

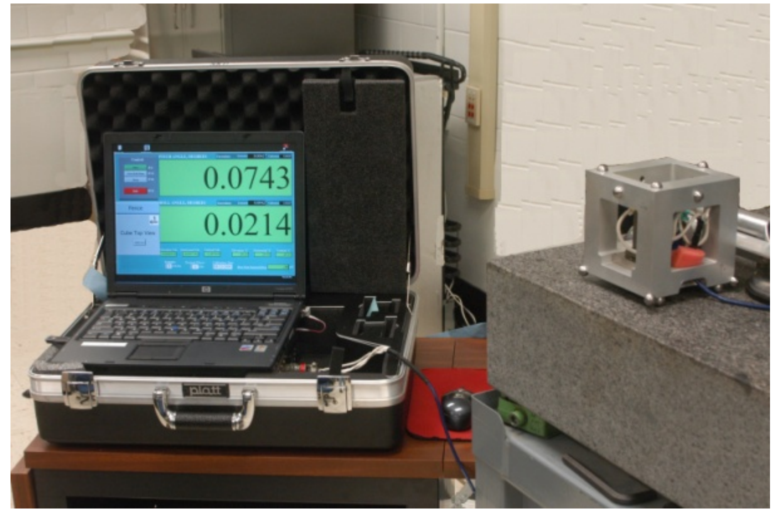
- **Calibration/Characterization Improvements**
  - **Implementation of statistical design and analysis techniques for baseline calibration**
  - **Model building based on:**
    - **Statistical significance**
    - **Variance inflation**
  - **Results:**
    - **Improved models (more robust)**
    - **New information on uncertainty intervals**
    - **Insight into instrument behavior and calibration experiments**

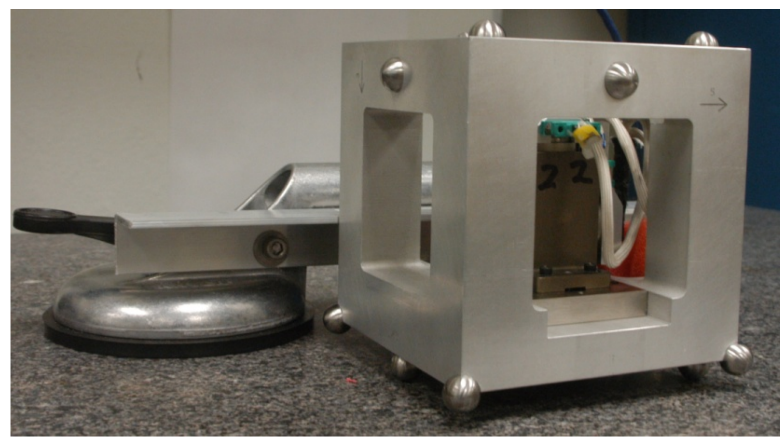


Baseline Calibration System

Angle Measurement Systems

- **Calibration/Characterization Improvements**
  - **Developed new calibration verification system – "Cube"**
    - **Adds statistical quality control**
  - **Results:**
    - **On-site, pre-test system performance verification**
    - **New information on uncertainty intervals**
    - **Insight into instrument behavior and calibration intervals**





Summary - Infusing Statistical Thinking

- **Developed strategic technical goals – One goal is to:**

- **Improve calibration/characterization and develop recommended practices – multi-component force and angle measurement systems lack traceable calibration system standards**

  - **Train SMEs on methodologies, tools and techniques**
  - **Bring statistical engineers into work hands-on projects to facilitate the training and improve knowledge transfer**

  - **Key: Continue questioning "WHY" – from the calibration systems to the experimental calibration designs to the calibration data analysis and model building**

Calibration Goal: *Produce an accurate mathematical model (matrix) to estimate the aerodynamic parameters from measured responses.*

Acknowledgements

- **Dr. Peter Parker**
- **Chris Lynn**
- **Hector Soto**
- **Tom Finley**
- **Jon Bader**
- **Mike Acheson**
- **Steve Helland**
- **Tim Marshall**